

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
29 September 2005 (29.09.2005)

PCT

(10) International Publication Number  
WO 2005/090559 A1

- (51) International Patent Classification<sup>7</sup>: C12N 9/12, 15/85
- (21) International Application Number:  
PCT/US2005/008977
- (22) International Filing Date: 18 March 2005 (18.03.2005)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/554,318 18 March 2004 (18.03.2004) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:  
US 60/554,318 (CIP)  
Filed on 18 March 2004 (18.03.2004)
- (71) Applicant (for all designated States except US): FRED HUTCHINSON CANCER RESEARCH CENTER [US/US]; 1100 Fairview Avenue North, P.O.Box 19024, Seattle, WA 98109-1024 (US).
- (72) Inventor; and  
(75) Inventor/Applicant (for US only): ROHRSCHEIDER, Larry, R. [—/US].
- (74) Agent: SHISHIMA, Gina, N.; FULBRIGHT & JAWORSKI, L.L.P., 600 Congress Avenue, Suite 2400, Austin, TX 78701 (US).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US (patent), UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),

[Continued on next page]

(54) Title: METHODS AND COMPOSITIONS INVOLVING S-SHIP PROMOTER REGIONS

-600 AGAAGTGATCTGGGCTGGACAGACTAGCTGAACTGGCCAGCTCTGGGTTC

-550 ATCAAGAAACCTTACCTCCATAACATAAAGTGTGATGGAGAAAGGCACCT

-500 AATGTC AACCTCAAACCCCTA <sup>Arb6</sup> CCTGCTGTGCACACACATACATCCACAC

-450 CAA

-400 <sup>Nkx3.1</sup> ATAAAT AAGTAAATAAATAAATA TTAGCTCTCCAGACCAAATCTTGGT

-350 GAAACC CATGCATT <sup>Oct4</sup> TTTGCTTTGTGTGTCTCTACAAACTGAAGGTAA

-300 GAAGCATGCTCCTTAGTAATTTATAGCAGTTTGCGTTTCCAGATTGAAA

-250 ACAGATTCTATAGGCTACACAGTGCTAAATGGATTATGCTCAGATACAGA

-200 TTGAAAAGGATACAGATTGAAAAGGGTCGGGGCTGGGCAGGATGACGG

-150 GCCAAC TATCTTTGCCCGGGCTT <sup>Gata / Lmo2</sup> TCTCCTCAGGGAAGGTTACAGGATT

-100 ACCACTGGGGTGTGGCCTATCTGCTGTTAGGACCTGAAATGCCTGGAGTG

-50 TTTCTA GTTCCCACTAGTTGTTGAACCTTACCTTGAACCTCTGCTCCAG

<sup>Initiator</sup> <sup>stem-SHIP region</sup>

(57) Abstract: The present invention concerns *s-ship* promoter and developmental decision promoter compositions and methods of using the promoter. It includes polynucleotides, vectors, host cells, and transgenic animal including a developmental decision promoter, for example, an *s-ship* promoter, controlling the expression of a heterologous nucleic acid. Methods of the invention concern methods of expressing a heterologous nucleic acid in a tissue-specific, developmental-specific, or temporally controlled manner. Other methods includes screening methods and therapeutic methods.

WO 2005/090559 A1



European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

**Published:**

— *with international search report*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

## DESCRIPTION

### METHODS AND COMPOSITIONS INVOLVING S-SHIP PROMOTER REGIONS

#### 5 BACKGROUND OF THE INVENTION

The present application claims priority to U.S. Provisional Patent Application 60/554,318, filed on March 18, 2004, which is hereby incorporated by reference in its entirety. The government may own rights in the invention pursuant to grant number CA82499  
10 from the National Institutes of Health.

##### 1. Field of the Invention

The present invention relates generally to the fields of molecular and developmental biology. More particularly, it concerns methods and compositions involving *s-SHIP* promoter  
15 regions, which can be used to promote transcription in particular cell types and at particular times during development.

##### 2. Description of Related Art

Stem cells have been the focus of tremendous interest in recent years because of the progress made in developmental and molecular biology and the promise of therapeutic  
20 applications in a wide variety of contexts, from heart disease to diabetes, and cancer to Parkinson's disease (see generally Abbott *et al.*, 2003; Daley, 2003; Hirai, 2002; Kondo *et al.*, 2003; Nakano, 2003). Toward fulfilling this promise, many researchers have engaged in extensive studies to characterize factors and pathways in stem cell development and to evaluate candidate therapeutic and diagnostic agents. Such agents include proteins that are  
25 gene products, sometimes heterologous, in the stem cells. The ability to express a transgene in stem cells is critical for providing data toward these endeavors. The study of genes normally expressed in stem cells has yielded not only information regarding the developmental, cellular, and molecular biology of these cells, but also useful tools for further studies.

30 Pathways involved in stem cell function include the protein phosphatidylinositol 3-kinase (PI3K), which becomes activated through cell surface receptors. PI3K is involved in the generation of phosphatidylinositol 3, 4, 5-triphosphate, which activates signaling

pathways leading to cell proliferation. The SH2-containing inositol 5'-phosphatase (SHIP1) removes the phosphate group from the D5 position of phosphatidylinositols, which is considered an significant feedback mechanism on cell activation for hematopoietic cells (Lioubin *et al.*, 1996; Rohrschneider *et al.*, 2000).

5        A form of SHIP1 lacking the SH2 domain has been identified and referred to as stem or short SHIP (s-SHIP) (Tu *et al.*, 2001). Tu *et al.* found the protein contains amino acids encoded by exons 6-27 of *SHIP1* and that it is expressed in embryonic and hematopoietic stem cells. It was initially unclear how s-SHIP protein was produced from the *ship1* gene. Kavanaugh *et al.* (1996) suggested that SIP-110 was a spliced product of SHIP1; however,  
10        Tu *et al.* (2001) proposed, based on its cDNA sequence, that it was transcribed from a promoter within the *SHIP1* gene. This was inferred from the fact that the first 44 nucleotides of the s-SHIP cDNA were at the 5' end immediately before exon 6 of *SHIP1*. These 44 nucleotides were not contained in the SHIP1 cDNA, but were identical to the 44 nucleotides of genomic *ship1* intron 5, immediately adjacent to exon 6. However, no functional evidence  
15        for an *s-ship* promoter was provided. Therefore, while a promoter with the tissue-specific expression of *s-ship* could be valuable from both a research and therapeutic/diagnostic perspective, further investigation of the *s-ship* gene was required to identify the promoter and characterize any tissue-specific activity.

Moreover, until now, a promoter providing the developmental-specific expression of  
20        the *s-ship* promoter, which includes expression in stem cells, has not previously been described. Thus, the present invention addresses these issues.

### SUMMARY OF THE INVENTION

The present invention concerns methods and compositions involving a functional and  
25        isolated *s-ship* promoter. The invention includes nucleic acid molecules, host cells, and transgenic organisms having an *s-ship* promoter, as well as methods of using the promoter for transcription, expression studies, stem cell analyses, and therapeutic applications. In addition the present invention relates to methods and compositions involving an isolated and function promoter capable of directing transcription in a) adult and embryonal stem/progenitor cells  
30        and in b) in adult and embryonal stem/progenitor cells that have differentiated to the point where the promoter directs transcription only when a developmental decision is required, *i.e.*, when the cell is in the resting or growing phase, or the transition from the resting to the

growing phase. This promoter will be referred to as a "developmental decision promoter." *S-ship* promoter regions discussed herein constitute a developmental decision promoter. Thus, it is contemplated that embodiments discussed with respect to an *s-ship* promoter can be applied more generally to a developmental decision promoter. Consequently, the present invention covers those embodiments with respect to development decision promoters.

The present invention concerns an *s-ship* promoter and its functional derivatives. The term "promoter" is used according to its ordinary and plain meaning to a person skilled in the art of eukaryotic transcriptional regulation. The terms "*s-ship* promoter" or "*s-ship1* promoter" refer to the nucleic acid sequence from the *s-ship* gene that is capable of promoting transcription of a nucleic acid sequence that is connected to it (downstream). Transcription can be assayed according to any number of ways known to those of skill in the art, including, but not limited to, an expression assay using a screenable or selectable marker; ribonuclease protection assay (RNAP), RT-PCR, and *in vitro* transcription reactions, all of which are well known to those of skill in the art and can be implemented using commercially available reagents and protocols (see generally, Sambrook *et al.*, 1989; Ausubel, 1992 and 1994, all of which are incorporated by reference).

Compositions of the invention include isolated polynucleotides comprising an *s-ship* promoter capable of promoting transcription. SEQ ID NO:1 is a 102 kb genomic mouse *ship1* sequence. In certain embodiments, the *s-ship1* promoter comprises, or has at least or at most 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 40, 50, 60, 70, 80, 90, 100, 110, 12, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200, 3300, 3400, 3500, 3600, 3700, 3800, 3900, 4000, 4100, 4200, 4300, 4400, 4500, 4600, 4700, 4800, 4900, 5000, 6000, 7000, 8000, 9000, 10000, 11000, 12000, 13000, 14000, 15000, 16000, 17000, 18000, 19000, 20000, 21000, 22000, 23000, 24000, 25000, 26000, 27000, 28000, 29000, 30000, or more contiguous nucleotides of the *ship1* gene, including SEQ ID NO:1-23, or any range derivable therein. In specific embodiments, *s-ship1* promoter includes one or more of the following regions of SEQ ID NO:1: from nucleotide (nt) 49485 to 61006 (11.5 kb-GFP construct);

49485 to 57111, which is 7626 nt (7.6 kb construct); from nt 49485 to 55810, which is 6326 nt (6.3 kb construct); from 54807 to 61006, but lacking 57109 to 57944 (6.2kb-GFP construct); from nt 51389 to 55810, which is 4421 nt (4.4 kb construct); from nt 52199 to 56423, which is 4224 nt (4.2 kb construct); from nt 53820 to 55810, which is 1990 nt (1.9 kb construct); from nt 54755 to 55810, which is 1055 nt (0.96 kb construct); or from nt 55668 to 55810, which is 142 nt (44 nt construct). It is further contemplated that the lengths of contiguous nucleotides discussed above can be applied with respect to these identified regions of SEQ ID NO:1, as well as any other sequence disclosed herein.

Moreover, any of these lengths or regions discussed in the context of SEQ ID NO:1, apply to the corresponding regions of SEQ ID NO:2, SEQ ID NO:3, and SEQ ID NO:4. SEQ ID NO:2 includes the genomic sequence for the mouse *ship* gene from exon 5 through exon 7 (exon 5, intron 5, exon 6, intron 6, and exon 7 inclusive). SEQ ID NO:3 is a mouse *s-ship* promoter sequence that includes the 560 nucleotides upstream of exon 6 (in intron 5). SEQ ID NO:4 is a human *s-ship* promoter sequence that includes the 560 nucleotides upstream of exon 6 (in intron 5). SEQ ID NO:5 is the mouse *s-ship* promoter region in the 11.5 kb-GFP construct. While these *s-ship* promoters provided are from human and mouse, the invention is not limited to these species. It is contemplated that mammalian *s-ship* promoters are contemplated, particularly those with homology to the sequences disclosed in the application. SEQ ID NO:6 is the sequence from intro 5 of *s-ship* that has a p53 family binding motif 5'-ATCTTTGCCC/GGGGCTTGTCCT-3', meaning that members of the p53 family of proteins have been shown to bind to sequences homologous or identical to this sequence. SEQ ID NO:7 is a sequence from the *s-ship* promoter that is homologous to a Pax8 binding sequence (5'-CACT/AGAAGGTT-3'). SEQ ID NO:8 is a sequence from the *s-ship* promoter that is homologous to a Smad 3/4 binding sequence (5'-GT/GC/GTGGGCCAG-3'). SEQ ID NO:9 is a sequence from the *s-ship* promoter that is homologous to a Stat 1/5 binding sequence (5'-TCAGGGA/GAG-3'). SEQ ID NO:10 is a sequence from the *s-ship* promoter that is homologous to a GATA/Lmo2 sequence (5'-GTGC/GCCTATCT-3'). It is understood by those of skill in the art that the convention of using a slash (/) indicates two alternative nucleotides at that position, where the slash separates the alternatives. It is also contemplated that sequences of the invention may include one or more of these consensus sequences for these binding sites. In certain embodiments, these motifs may be repeated singly or in combination with one another.

Moreover, it is contemplated that the promoter contains enhancer activity. In some embodiments of the invention, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 12, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600 contiguous nucleotides, or any range derivable therein, from the region including the first 600 nucleotides upstream of exon 6 are included in *s-ship* promoters of the invention. It is also contemplated that segments that are at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identical to any of these regions or capable of hybridizing to the complement of such a regions are contemplated as part of the invention. Such segments may further comprise sequences in intron 6, exon 5, exon 6 or other regions of the *ship1* gene.

It is contemplated that functional derivatives of the *s-ship* promoter also contemplated by the invention. Functional derivatives of an *s-ship* promoter may be at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identical to the polynucleotides discussed herein. Such derivatives may also be characterized by any of the lengths of contiguous nucleotides discussed above. Moreover, polynucleotides of the invention include those that are capable of hybridizing to all or part of the recited lengths of SEQ ID NOs:1-10 discussed herein, including those particular regions recited in the previous paragraph. Such polynucleotides may be capable of hybridization under high, medium or low stringency conditions.

In specific embodiments, the *s-ship* promoter is capable of promoting tissue-specific transcription. Transcription may be accomplished, in some embodiments of the invention, in skin, a hair follicle, cornea, embryo, gonads, mammary gland, pancreas, and/or vascular smooth muscle. It is also contemplated that transcription may be achieved in cells qualified as or with characteristics of stem cells, which may or may not be derived from skin, a hair follicle, cornea, embryo, gonads, mammary gland, pancreas, and/or smooth muscle. In some embodiments, transcription is achieved in a hematopoietic cell in a tissue-specific manner, including in hematopoietic stem or progenitor cells, but also in more mature or differentiated cells.

It is specifically contemplated that the *s-ship* promoter directs transcription in a developmentally and/or cell-cycle-dependent manner. In some embodiments, the *s-ship* promoter directs transcription in stem or progenitor cell that differentiates to a point so that the promoter no longer provides expression at all or provides expression during certain times, such as when it is preparing for a growth and/or developmental phase. As discussed above,

the invention concerns developmental decision promoters such as *s-ship* and embodiments discussed with respect to *s-ship* can be applied respect to developmental decision promoters, and vice versa.

In some embodiments, the present invention includes a promoter that is capable of directing transcription in cells that qualify as stem or progenitor cells and/or cells that have undergone some differentiation but are not terminally differentiated and that are not in a resting state. In some embodiments, the invention provides isolated poly nucleotides, expression cassettes, vectors and host cells comprising a heterologous nucleic acid sequence under the control of a developmental decision promoter.

In these embodiments, the developmental decision promoter is capable of providing expression in embryonic stem cells. In other embodiments, the promoter is capable of providing expression in adult stem cells. It is contemplated that the adult stem cells are differentiated but not terminally differentiated; in other words, they are self-renewing and capable of being differentiated into other cell types derivative of the stem cell. For example, the adult stem cell may be a hematopoietic or epidermal stem cell meaning it is capable of self-renewing and becoming any hematopoietic or epidermal skin cell. The term "terminally differentiated" is used according to its ordinary and plain meaning according to those of ordinary skill in the art.

In other embodiments, the developmental decision promoter is capable of providing expression in adult stem cells that are in growing phase (*i.e.*, in a non-resting phase of mitosis or meiosis). In certain other cases, the promoter is capable of providing expression in a cell from mouse embryonic development stages E3-E18.5. The E numbers refer to age of the embryo, based on days, from approximate conception, for example, as set forth on the World Wide Web at the following address:

genex.hgu.mrc.ac.uk/CDROM\_online/macd/html\_shdw\_links/mastaging.html. It is contemplated the expression may be achieved at mouse embryonic development stage E1, E2, E3, E3.5, E4, E4.5, E5, E5.5, E6, E6.5, E7, E7.5, E8, E8.5, E9, E9.5, E10, E10.5, E11, E11.5, E12, E12.5, E13, E13.5, E14, E14.5, E15, E15.5, E16, E16.5, E17, E17.5, E18, E18.5, E19, E19.5, E20 or later, or any combination derivable therein, or any corresponding stage of development in another species of mammal. In some embodiments, the promoter can provide expression throughout all these stages of development (constitutive) or through a subset (not constitutive throughout). In particular embodiments, the promoter is capable of directing transcription in stem/progenitor cells of an embryo and also in adult stem cells periodically (non-constitutively) or constitutively.



It is further contemplated that the developmental decision promoter is further capable of providing expression in a cell that is in a developed animal. In other words, a developed animal refers to an animal that has already been born and is no longer an fetus or embryo. Thus, the cell may be in an animal that has already been born and it may be near or  
5 surrounded by differentiated cells or tissue. For example, the cell may be a stem or progenitor cell in the developed animal.

In particular embodiments, the developmental decision promoter is an *s-ship* promoter region comprises a sequence that can hybridize under stringent conditions to nucleic acid segment comprising the complement of i) at least 20 contiguous nucleic acids of SEQ ID  
10 NO:1, SEQ ID NO:2, or SEQ ID NO:3; or ii) SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9 and/or SEQ ID NO:10.

In some embodiments of the invention, an *s-ship* promoter is operably attached to a heterologous nucleic acid. The term "heterologous" is used according to its plain and ordinary meaning to a person skilled in the art of molecular biology. It is a relative term and  
15 in the context of an *s-ship* promoter, it refers to a nucleic acid sequence that is not normally found in nature (with respect to sequence and position) with the *s-ship* promoter. In other words, it refers to any nucleic acid that is not the entire genomic sequence of the *s-ship* gene. In some embodiments, the *s-ship* promoter is connected to a nucleic acid sequence encoding part of the *s-ship* gene product or all or part of an *s-ship1* cDNA sequence. Alternatively, the  
20 *s-ship* promoter may be placed at a location different than is found in nature.

Because recombinant cells and transgenic animals, including knockout versions thereof, are part of the invention, the present invention further encompasses nucleic acids containing an *s-ship* gene or a portion thereof and a marker sequence, wherein the *s-ship* gene is disrupted by the marker sequence. In some embodiments, the nucleic acid is under the  
25 control of a promoter, which is an *s-ship* promoter in further embodiments. The promoter may also be constitutive, inducible, or conditional. Promoters discussed herein may be tissue-specific (spatially restricted), developmental-specific (providing transcription at specific developmental stages or times), and/or temporally restricted.

The present invention further concerns expression cassettes, vectors, and host cells  
30 that contain or include polynucleotides having an *s-ship* promoter that has been isolated away from its chromosomal context. The polynucleotides and embodiments discussed above may be implemented with respect to expression cassettes, vectors, and host cells.

It is contemplated that the *s-ship* promoter may control the transcription of a nucleic acid sequence encoding a marker. In some embodiments, the marker is colorimetric,

enzymatic, or fluorescent. Examples include, but are not limited to,  $\beta$ -galactosidase, chloramphenicol acetylase, luciferase, and green fluorescent protein. In further embodiments, a heterologous nucleic acid segment encodes a therapeutic or diagnostic gene product. The therapeutic or diagnostic gene product may be a protein or RNA molecule, such as an siRNA or miRNA molecule. In some embodiments, the therapeutic gene product is selected from the group consisting of a tumor suppressor, an oncogene, a cytokine, a cytokine receptor, a differentiation-inducer, growth factor, and a growth factor receptor. It is contemplated that more than one heterologous sequence or gene may be placed under the control of a promoter. Examples of such proteins are well known to those of skill in the art, and include, but are not limited to, interleukins (IL-2, -6, -8, -9, -10, -11, -12, -13, -14, -15, -16, -17, -18, -19, -20, -21, -22, -23, -24, etc.), interferons, receptor tyrosine kinases and their ligands (kit/steel, CSFR/CSF, GM-CSFR/GM-CSF, PDGFR/PDGF, flk-1/VEGF, Lif, EGF, FGF, etc.), transforming growth factors  $\alpha$  and  $\beta$ , Epo, IGF, tumor necrosis factor  $\alpha$  and  $\beta$ . A number of examples can be seen on the world wide web at [indstate.edu/thcme/mwking/growth-factors.html](http://indstate.edu/thcme/mwking/growth-factors.html), which is specifically incorporated by reference. In specific embodiments, it is contemplated that the heterologous encoded protein can transform or immortalize a cell, such as an oncogene. In certain embodiments, a stem cell can be immortalized or transformed.

In some embodiments of the invention, a vector is a plasmid, YAC, BAC, or virus. Viruses include adenovirus, adeno-associated virus, retrovirus, flavivirus, and vaccinia virus.

Compositions of the invention may be prepared in a pharmaceutically or pharmacologically acceptable formulation. Such formulations are well known to those of skill in the art for use in *in vivo* contexts.

Other aspects of the invention include host cell having an *s-ship* promoter operably attached to a heterologous nucleic acid segment. In some embodiments, the host cell is eukaryotic, though it may be prokaryotic. In specific embodiments, the host cell is from a mammal, insect, bacteria, or yeast. Cells from monkeys, mice, rats, rabbits, hamsters, ferrets, and humans are specifically contemplated for use with nucleic acids of the invention. In some cases, the host cell is an embryonic cell, which may specifically be a blastocyst cell. In other cases, the host cell is a stem or progenitor cell. In some cases, the cell is a hematopoietic cell, meaning any cell in that lineage. It is contemplated that the cell may be *in vitro* or *in vivo*.

Cells that can be used according to methods and compositions of the invention include, but are not limited to, CD34+ cells (cells expressing CD34 on their surface),

undifferentiated cells, stem cells, progenitor cells, cord blood cells, placental cells, neonatal or fetal cells, immature cells, pluripotent cells, and totipotent cells. The term "stem cell" is used according to its ordinary meaning, for example, as described by the National Institutes of Health (on the World Wide Web at [stemcells.nih.gov](http://stemcells.nih.gov)). Stem cells 1) are "capable of dividing and renewing themselves for long periods"; 2) are unspecialized; and, 3) can give rise to specialized cell types.

The invention specifically contemplates the use of embryonic stem cells, adult stem cells, or neonatal and fetal stem cells. An adult stem cell typically refers to a stem cell from a particular organ or tissue that is capable of differentiating into one or more cells of that organ or tissue. Umbilical cord blood contains stem cells that are similar to embryonic stem cells in that they are believed to be capable of being differentiated into a number of different cell types, as opposed to cell types of one particular organ or tissue. Umbilical cord blood refers to blood that remains in the umbilical cord and placenta following birth and after the cord is cut. "Placental blood" is understood to be synonymous with cord blood; similarly, cord blood stem cell is considered synonymous with placental or placental blood stem cell. The use of stem cells from umbilical cord blood is specifically contemplated in certain embodiments of the invention. In some but not all cases, the use of other stem cells is specifically not considered part of the invention, particularly the use of pancreatic/endocrine progenitor or stem cells is not considered for use with some embodiments. Furthermore, cells of the invention may be characterized by cell surface antigens. Cell surface antigens and their correlation with cell type and cell development are known to those of ordinary skill in the art.

It will be understood that cultures or samples containing cells discussed above are also contemplated for use according to methods and compositions of the invention.

Further embodiments of the invention include cells for use in the generation of transgenic organisms (knock-in and knock-out). Accordingly, there are recombinant host cells in which one or both *s-ship* genes is disrupted by marker sequence or in which all or part of an *s-ship* gene is flanked by an excisable sequence, such as a loxP sequence. The marker sequence serves the purpose of showing when the transgenic sequence is present or absent in the cell.

The present invention further concerns transgenic animals comprising an *s-ship* promoter operably attached to a heterologous nucleic acid segment. Mammals are specifically contemplated, particularly mice. In some cases, the invention involves a mammal having cells comprising an *s-ship* transgenic sequence. The *s-ship* sequence may be knocked in or out in a restricted or controlled manner. For example, whether it is knocked in or out may be

controlled in a tissue-specific, inducible, conditional, developmental or temporal manner. Consequently, animals may have heterologous genes under the control of a promoter or system that operates in that way. The Cre-Lox system is one example. The transgene of interest itself may not be under the control of a limited promoter, but a secondary gene whose product initiates the knock-in or knock-out process may be under such a promoter. In one embodiment, animals of the invention may have an *s-ship* transgenic sequence that includes an *s-ship* coding sequence flanked by loxP sequences. They may also have a heterologous nucleic acid sequence encoding a Cre recombinase. In some cases, the nucleic acid sequence encoding the Cre recombinase is under the control of an inducible or conditional promoter. Transgenic animals of the invention are not limited by the Cre-Lox system, which serves as an example of how expression may be controlled.

A number of methods are included as part of the present invention. In some embodiments, there are methods for expressing a recombinant nucleic acid in a cell comprising: a) transfecting the cell with an expression cassette comprising an *s-ship* promoter operably attached to the recombinant nucleic acid, wherein the nucleic acid is transcribed. The cell may be any of the host cells discussed above. Moreover, it is contemplated that embodiments may be carried out with a developmental decision promoter, which may be an *s-ship* promoter region. In some embodiments there are methods for expressing a nucleic acid in a stem cell comprising providing to a cell a polynucleotide including the nucleic acid under the control of a developmental decision promoter, wherein the nucleic acid is expressed in the cell. It is contemplated that a cell may be in a subject. The cell may have been provided with a nucleic acid *in vivo* or *in vitro*. In the latter case, a cell may be introduced into a subject thereafter.

Alternatively, an isolated nucleic acid encoding a developmental decision promoter can be provided to cell such that the promoter integrates into the cell's genome to drive expression of a gene that becomes operably linked to the promoter. The present invention covers methods and compositions for implementing the expression strategy.

Other embodiments of the invention concern methods of screening for a candidate substance that regulates activity of the *s-ship* promoter comprising a step selected from the group consisting of: (a) contacting a nucleic acid comprising an *s-ship* promoter with an *s-ship* promoter binding protein and the candidate substance under conditions that allow binding between the protein and the promoter and determining whether the candidate compound modulates the binding between the protein and the promoter; and (b) contacting the candidate substance with a cell comprising the *s-ship* promoter operably attached to a

reporter gene coding for an expression product and assaying for expression of the reporter gene expression product. One or both steps may be employed. Ways of determining whether the candidate compound modulates binding between a protein and the promoter are well known to those of skill in the art. The compound may inhibit, reduce, decrease, eliminate, increase, promote, tighten the binding between the protein and the promoter. Assays for such an interaction include, but are not limited to, electrophoretic mobility shift assays (EMSA), DNA footprinting, functional transcription assays—as described above—Southwestern assays, and PCR-based assays.

The present invention also includes methods for identifying stem cells in a population of cells comprising: (a) administering to cells in the population a nucleic acid comprising an *s-ship* promoter operably attached to a reporter or marker gene. The reporter or marker gene is then used to identify positively-expressing cells, which would indicate the cell is a stem cell. The cell may be in an organ and/or in an animal. In some embodiments, methods include sorting cells based on expression of the reporter or marker gene. In addition to the assays discussed above, FACS analysis may be employed, in addition to other cell sorting techniques. Methods include differentiation of the cells.

Aspects of the invention also concern methods for screening for a modulator of cell function comprising: a) transfecting a stem or hematopoietic cell with an expression cassette comprising an *s-ship* promoter operably attached to a nucleic acid encoding a candidate modulator; and, b) assaying the cell for a cell function, wherein a difference in cell function in the cell as compared to a cell in the absence of the candidate modulator is indicative of a modulator. The term “modulator” refers to a substance that affects cell function. It may affect cell function by acting on or through a pathway. The modulator may inhibit, reduce, eliminate, decrease, increase, promote, induce, or enhance a particular cell function or result of a pathway in the cell. It is contemplated that this method may be employed to identify a modulator as a candidate therapeutic agent for the treatment of a blood-related disease or condition.

Therapeutic methods are also provided by the present invention. Methods are not necessarily limited to a particular disease or condition. It is contemplated that any method in which expression in stem cells or cells in which the *s-ship* promoter can function are contemplated for use in therapeutic methods of the invention. For example, the method may be applied to pancreatic disorders and diseases.

Thus, in some embodiments, there is a method of treating a patient with a blood-related disease or condition comprising: a) transfecting a cell with an expression cassette

comprising an *s-ship* promoter region operably attached to a therapeutic nucleic acid; and, b) administering the cell to the patient. Blood-related disease or condition include blood-related cancers—such as leukemia, lymphoma, or myeloma—and anemia. In some cases, the blood-related condition can be treated using stem cell replacement therapy.

5 Other methods of the invention include providing a method of treating a tumor comprising providing to stem cells of the tumor an agent that promotes their destruction. In some embodiments a patient with a tumor is provided with a host cell or expression construct comprising a developmental decision promoter such as an *s-ship* promoter region that provides expression for a therapeutic agent in the tumor stem cell. The therapeutic agent may  
10 a protein or a nucleic acid. In certain embodiments, the agent promotes apoptosis or cell death of the tumor stem cell, such as with a toxin or apoptosis inducer.

Other methods of the invention include ways of tracking stem cells or isolating stem cells. Expression using developmental decision promoters can be used to track or isolate stem cells by virtue of their expressing a product that can be tracked or used to isolate the stem  
15 cells. In terms of tracking, the product may be some kind of reporter, which may or may not be a cell surface marker. In the case of isolating stem cells, the product will allow the stem cells to be separated from non-stem cells, such as by FACS analysis based on an expressed cell surface marker.

Cells for therapeutic use may, in addition to the cells discussed above, be bone  
20 marrow cells, or be autologous or allogeneic.

It is contemplated that any embodiment discussed with respect to any method or composition described herein can be implemented with respect to any other method or composition described herein. For example, an embodiment discussed with respect to an *s-ship* promoter region applies to a developmental decision promoter, and vice versa. Similarly,  
25 an embodiment discussed with a polynucleotide, primer, expression constructs, host cells, transgenic organisms, and method of the invention are contemplated for use with any other polynucleotides, primers, expression constructs, host cells, transgenic organisms, and methods of the invention, and vice versa.

The use of the term “or” in the claims is used to mean “and/or” unless explicitly  
30 indicated to refer to alternatives only or the alternatives are mutually exclusive, although the disclosure supports a definition that refers to only alternatives and “and/or.”

Throughout this application, the term “about” is used to indicate that a value includes the standard deviation of error for the device or method being employed to determine the value.

5 Following long-standing patent law convention, the words “a” and “an,” when used in conjunction with the word “comprising” in the claims or specification, denotes one or more, unless specifically noted.

10 Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples, while indicating specific embodiments of the invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

15 The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

20 **FIG 1.** *ship1* genomic segments cloned into a promoter-less expression vector for testing cell-specific promoter activity. The upper line represents the general *ship1* genomic region containing potential activity for cell-specific s-SHIP expression. Intron 5 contains the likely promoter activity and transcription is proposed to begin before exon 6. The 44-intronic nucleotides, contained in the s-SHIP cDNA, are shown as red. A 7.6 kb genomic fragment  
25 (second line down), as well as the indicated sub-fragments, were cloned into a promoter-less plasmid for GFP expression. The design and construction of the plasmid is detailed in Materials and Methods.

**FIG. 2.** Flow cytometry analysis of cell type-specific promoter activity in D3 ES cells vs. NIH3T3 cells. Each construct shown was cloned into a promoter-less GFP plasmid,  
30 which was linearized and electroporated into D3 ES cells, or transfected into NIH3T3 cells. G418 resistant colonies were then examined by flow cytometry for GFP expression. Two different “empty vector” negative controls were utilized depending on whether the insert

contained a splice acceptor or both splice acceptor and donor. Both these plasmids without genomic insert were negative for GFP expression in both cell types, but only a single negative-control vector is shown. Two positive-control plasmids were utilized in each experiment. These controls expressed GFP from an IRES, and one expressed a protein insert, both were positive in each cell type.

**FIG. 3A-B.** Structure of the 11.5-kb and 6.2-kb transgenic promoter-GFP constructs for in vivo analysis. **FIG. 3A.** Two promoter transgenic constructs were prepared. The first construct is called the 11.5kb-GFP transgene, and contains the entire genomic *ship1* segment from the Sac I site near the 5' end of intron 5 through the putative translation start site at an ATG preceded by a suitable Kozac sequence within exon 7. The translational start ATG for the enhanced GFP is fused, in frame, to the likely ATG translational start for s-SHIP. A second transgenic construct, called the 6.2kb-GFP transgene, is identical to the 11.5kb-GFP construct, except it contains only 0.96 kb upstream of exon 6, and lacks 833 nt within intron 6. **FIG. 3B.** Transgenic copy numbers were estimated by semi-quantitative PCR analysis relative to the endogenous diploid *gab2* gene.

**FIG. 4.** Computer analysis of 600 nucleotides of the intron-5 transgene promoter region. A. The region immediately upstream of exon 6 is shown with potential transcription factor binding motifs determined by analysis using the MatInspector program. Only the factors with matrix and core similarity greater than 0.9 are shown. Those factor motifs within the strand shown are over-lined, while those factors potentially interacting with the complementary strand are shown underlined. The SSR or stem-SHIP region identified by Tu *et al.*, 2001 is in bold, and an initiator sequence for transcription is situated at the beginning of the SSR. Exon 6 (not shown) begins at the 3' end of the SSR.

**FIG. 5.** The two primary proteins, s-SHIP and SHIP1, are produced from the *ship1* gene. The domain structure of the two proteins is shown above the *ship1* genomic intron/exon organization. Transcription for the full-length 145 kDa SHIP1 protein initiates in promoter 1 (Pro1), utilizes all 27 exons, and translation begins in the exon 1 encoded region. The stop codon is the first three nucleotides of exon 27. Transcription for the s-SHIP protein begins within intron 5 (Pro 2), and downstream is identical to the SHIP1 product. Translation, however, presumably begins in the first ATG of exon 7. Both transcripts and protein sequences are therefore identical from the ATG in exon 7 through the stop codon in exon 27. The dashed lines indicate translation start and stop points for each protein within the genomic exons.



**FIG. 6.** The 560-nucleotide regions immediately upstream of exon 6 from the mouse and human sequences were compared. Inr indicates the initiator sequence. Binding sites are also identified.

**FIG. 7.** p53 binding sequences with half sites are depicted. This sequence is SEQ ID  
5 NO:7.

### **DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS**

The present invention is based on the isolation and characterization of the *s-ship1* promoter, which can be used to promote transcription. Methods and compositions involving the *s-ship1* promoter are provided herein. In some embodiments, they take advantage of the  
10 tissue specificity of *s-ship1* expression. *s-ship1* encodes a protein whose expression has been observed in limited cell populations, and thus, the tissue-specificity of its promoter can be exploited in a number of different ways.

#### **I. SHIP1 and s-SHIP Background**

The *s-ship1* promoter was studied because of the function and expression patterns for  
15 the *s-ship1* (also referred to as *s-ship*) and *ship1* gene products. The murine SHIP1 protein is encoded in 27 exons of the *Inpp5d* (inositol polyphosphate-5-phosphatase D) locus, spanning approximately 102 kbps on chromosome 1 at position 57.0 cM of the genetic map, or cytoband C5 of the cytogenetic map (reviewed in Rohrschneider *et al.*, 2000; Wolf *et al.*, 2001; NCBI databases). The full-length protein is 145 kDa, but splicing, involving exons 25  
20 and 26, can produce 4 additional proteins ranging in size from 109-135 kDa (Lucas and Rohrschneider, 1999; Wolf *et al.*, 2000). These splicing reactions affect the 350-amino acid C-terminal tail region and its numerous protein interaction motifs, such as those binding PTB, SH2, and SH3 domains.

The prominent structural features of the SHIP1 protein dictate its major functional  
25 interactions. The SHIP1 SH2 domain has general specificity for tyrosine-phosphorylated Yxx(L/I/V) amino acid motifs, and its inositol 5'phosphatase domain removes phosphate from the 5' position of inositol(3,4,5)P<sub>3</sub>, phosphatidylinositol(3,4,5)P<sub>3</sub> or Inositol(1,3,4,5)P<sub>4</sub> [see Sly *et al.*, (2003) for review]. The tyrosine-phosphorylated C-terminal tail interacts directly with the PTB domain of Shc and Dok proteins (Lioubin *et al.*, 1996; Sattler *et al.*,  
30 2001; Tamir *et al.*, 2000), and a potential interaction motif for the SH2 domain of the p85 component of the p85/PI3K is present in the full-length SHIP1 (Gupta *et al.*, 1999; Lucas and Rohrschneider 1999), but eliminated by the splicing events producing the  $\alpha$  and  $\beta$  isoforms

(Rohrschneider *et al.*, 2000). Polyproline-rich interaction motifs for the SH3 domains of Grb2 also are present in the C-tail region (Kavanaugh *et al.*, 1996). The SHIP1 proteins (*e.g.*, the 145 kDa protein and isoforms thereof) are expressed in hematopoietic cells and testes, with lower expression observed in a few other adult tissues (Q. Liu *et al.*, 1998, reviewed in Rohrschneider, 2003).

Functionally, both biochemical and genetic experiments indicate SHIP1 is a negative regulator of myeloid cell proliferation, survival, and perhaps chemotaxis (see Sly *et al.*, 2003; Rohrschneider, 2003). Also, SHIP1 negatively regulates degranulation, inflammatory cytokine release, and adhesion for mast cells, and SHIP1 is a component of negative signaling (anergy) in B cell proliferation. The molecular mechanisms for most of these effects require the attachment of the SHIP1 SH2 domain to the cytoplasmic portions of transmembrane receptors containing appropriate tyrosine-phosphorylated interaction motifs. There, the SHIP1 inositol-5'phosphatase domain converts the plasma membrane PI3K-produced substrate, phosphatidylinositol(3,4,5)P<sub>3</sub> to phosphatidylinositol(3,4)P<sub>2</sub> effectively terminating proliferation signals. Therefore, the SH2 domain of SHIP1 plays a critical role in initiating many of these negative biological effects.

An additional smaller protein from the *ship1* locus is described as an SH2-less 104 kDa protein (Tu *et al.*, 2001). This product is called s-SHIP, with the prefix signifying its only known expression within two stem cell types (*i.e.*, ES cells and lineage-depleted Sca1-positive cells of the bone marrow). This protein was first described by Kavanaugh *et al.* (1996) and called SIP-110 in the human; but details of its existence were unclear until Tu *et al.* (2001) defined the cDNA and demonstrated endogenous expression in the two cell types described above. Thus, structurally, s-SHIP differs from SHIP1 only by the lack of the N-terminal SH2 domain; but biochemically, s-SHIP also lacks tyrosine phosphorylation and association with Shc (Kavanaugh *et al.*, 1996; Tu *et al.*, 2001). Nevertheless, s-SHIP constitutively interacts with Grb2. The lack of an SH2 domain in s-SHIP indicates its interaction mechanism with target proteins probably differs from that of SHIP1; however, the biological functions of s-SHIP are not known.

## II. Nucleic Acids

### A. Polynucleotides

The *s-ship1* promoter was identified as a strong promoter for s-SHIP by analyses of the genomic *ship1* intron-5 region in driving GFP expression both *in vitro* and *in vivo*. This

promoter exhibited cell-type specific expression in ES cells, and mice transgenic for the promoter (the 11.5kb-GFP transgene) showed tissue-specific GFP expression within the inner cell mass of the blastocyst. Transgenic mice produced with a shorter promoter construct (the 6.2kb-GFP transgene) expressed GFP throughout the blastocyst, suggesting the absence of  
5 negative regulatory regions in the shorted transgene. RT-PCR analysis demonstrated *s-SHIP* expression within the blastocyst. These results indicate that the 11.5-kb promoter region of the transgene contains the information for tissue-specific expression of *s-SHIP*, as well as tissue-specific shut-off of this protein. It is specifically contemplated that this promoter and the transgenic mice will be useful for future examination of GFP-expression in potential  
10 stem/progenitor cells of the embryo and the adult mouse.

The present invention concerns polynucleotides, isolatable from cells, that are free from total genomic DNA and that contain an *s-ship* promoter. It is contemplated that the *s-ship1* promoter is capable of directing transcription of nucleic acid sequence. Transcription may be directed in a tissue-specific or developmental manner. The nucleic acid sequence may  
15 encode a peptide or polypeptide, or it may also encode an RNA molecule that is not translated into a protein.

A "promoter" is a control sequence that is a region of a nucleic acid sequence at which initiation and rate of transcription are controlled. It may contain genetic elements at which regulatory proteins and molecules may bind, such as RNA polymerase and  
20 transcription factors. The phrases "operatively positioned," "operatively linked," "under control," "operatively attached," and "under transcriptional control" mean that a promoter is in a correct functional location and/or orientation in relation to a nucleic acid sequence to control transcriptional initiation and/or expression of that sequence. Typically, the promoter is located 5' or upstream from the strand of sequence to be transcribed. A promoter may or  
25 may not be used in conjunction with an "enhancer," which refers to a cis-acting regulatory nucleic acid sequence involved in the transcriptional activation of a nucleic acid sequence. More particularly, it refers to a nucleic acid sequence that is tissue-specific and stimulates transcription regardless of orientation (forward or reverse orientations both work). The inventors believe that within the first 600 nucleotides upstream of exon 6 there is enhancer  
30 activity. Consequently, it is contemplated that all or part of this region may be included in nucleic acid construct containing segment of SEQ ID NO:1.

As used herein, the term "DNA segment" or "nucleic acid segment" refers to a DNA or nucleic acid molecule that has been isolated free of total genomic DNA of a particular

species. Therefore, a DNA segment encoding a polypeptide refers to a segment that contains wild-type, polymorphic, or mutant polypeptide-coding sequences yet is isolated away from, or purified free from, total mammalian or human genomic DNA. Included within the term "DNA segment" are a polypeptide or polypeptides, DNA segments smaller than a polypeptide, and recombinant vectors, including, for example, plasmids, cosmids, phage, viruses, and the like.

As used in this application, the term "s-ship polynucleotide" refers to an s-ship-encoding nucleic acid molecule. The term "cDNA" is intended to refer to DNA prepared using messenger RNA (mRNA) as template.

It also is contemplated that a particular polypeptide from a given species may be represented by natural variants that have slightly different nucleic acid sequences but, nonetheless, encode the same protein.

Similarly, a polynucleotide comprising an isolated or purified wild-type, polymorphic, or mutant polypeptide gene refers to a DNA segment including wild-type, polymorphic, or mutant polypeptide coding sequences isolated substantially away from other naturally occurring genes or protein encoding sequences. In this respect, the term "gene" is used for simplicity to refer to a functional protein, polypeptide, or peptide-encoding unit. As will be understood by those in the art, this functional term includes genomic sequences, cDNA sequences, and smaller engineered gene segments that express, or may be adapted to express, proteins, polypeptides, domains, peptides, fusion proteins, and mutants. A nucleic acid encoding all or part of a native or modified polypeptide may contain a contiguous nucleic acid sequence encoding all or a portion of such a polypeptide of the following lengths: about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 441, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000, 1010, 1020, 1030, 1040, 1050, 1060, 1070, 1080, 1090, 1095, 1100, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 9000, 10000, or more nucleotides, nucleosides, or base pairs.

In particular embodiments, the invention concerns isolated DNA segments and recombinant vectors incorporating an *s-ship* promoter with a heterologous nucleic acid sequence or a ship or s-ship cDNA segment. Thus, an isolated DNA segment or vector containing a DNA segment may encode, for example, the heterologous nucleic acid

sequence. The term "recombinant" may be used in conjunction with a polypeptide, the name of a specific polypeptide, a nucleic acid sequence, or a host cell, and this generally means that the entity involves or involved a nucleic acid molecule that was manipulated *in vitro* using recombinant DNA technology.

5       The nucleic acid segments used in the present invention, regardless of the length of the coding sequence itself, may be combined with other nucleic acid sequences, such as enhancers, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed,  
10       with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol.

      It is contemplated that the nucleic acid constructs of the present invention may encode full-length polypeptide from any source or encode a truncated version of the polypeptide, such that the transcript of the coding region represents the truncated version. The truncated  
15       transcript may then be translated into a truncated protein. Alternatively, a nucleic acid sequence may encode a full-length polypeptide sequence with additional heterologous coding sequences, for example to allow for purification of the polypeptide, transport, secretion, post-translational modification, or for therapeutic benefits such as targetting or efficacy. As discussed above, a tag or other heterologous polypeptide may be added to the modified  
20       polypeptide-encoding sequence, wherein "heterologous" refers to a polypeptide that is not the same as the modified polypeptide.

      In a non-limiting example, one or more nucleic acid constructs may be prepared that include a contiguous stretch of nucleotides of sequences disclosed herein, including the *s-ship* promoter.

25       A nucleic acid construct may be at least 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 15,000, 20,000, 30,000, 50,000, 100,000, 250,000, 500,000, 750,000, to at least 1,000,000 nucleotides in length, as well as constructs of greater size, up to and including chromosomal sizes (including all intermediate  
30       lengths and intermediate ranges), given the advent of nucleic acids constructs such as a yeast artificial chromosome are known to those of ordinary skill in the art. It will be readily understood that "intermediate lengths" and "intermediate ranges," as used herein, means any length or range including or between the quoted values (*i.e.*, all integers including and between such values).

It is specifically contemplated that nucleic acids of the invention may include, be at most, or be at least 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200, 3300, 3400, 3500, 3600, 3700, 3800, 3900, 4000, 4100, 4200, 4300, 4400, 4500, 4600, 4700, 4800, 4900, 5000, 5100, 5200, 5300, 5400, 5500, 5600, 5700, 5800, 5900, 6000, 6100, 6200, 6300, 6400, 6500, 6600, 6700, 6800, 6900, 7000, 7100, 7200, 7300, 7400, 7500, 7600, 7700, 7800, 7900, 8000, 8100, 8200, 8300, 8400, 8500, 8600, 8700, 8800, 8900, 9000, 9100, 9200, 9300, 9400, 9500, 9600, 9700, 9800, 9900, 10000, 10100, 10200, 10300, 10400, 10500, 10600, 10700, 10800, 10900, 11000, 11100, 11200, 11300, 11400, 11500, 11600, 11700, 11800, 11900, 12000 or more contiguous nucleotides (or any range derivable therein) of nucleic acid disclosed in this application, including, but not limited to SEQ ID NO:1, intron 5 of the mouse *s-ship* gene, an *s-ship* promoter, and any other SEQ ID NOs such as SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or any of SEQ ID NO:6-11 and/or 12-25.

The various probes and primers designed around the nucleotide sequences of the present invention may be of any length. By assigning numeric values to a sequence, for example, the first residue is 1, the second residue is 2, *etc.*, an algorithm defining all primers can be proposed:

$$n \text{ to } n + y$$

where  $n$  is an integer from 1 to the last number of the sequence and  $y$  is the length of the primer minus one, where  $n + y$  does not exceed the last number of the sequence. Thus, for a 10-mer, the probes correspond to bases 1 to 10, 2 to 11, 3 to 12 ... and so on. For a 15-mer, the probes correspond to bases 1 to 15, 2 to 16, 3 to 17 ... and so on. For a 20-mer, the probes correspond to bases 1 to 20, 2 to 21, 3 to 22 ... and so on.

It also will be understood that this invention is not limited to the particular nucleic acid sequences of SEQ ID NO:1. Recombinant vectors and isolated DNA segments may

therefore variously include coding regions, coding regions bearing selected alterations or modifications in the basic coding region, or they may encode biologically functional equivalent sequences. For example, mutations can be made to SEQ ID NO:1-25 that potentially enhance or alter function relative to the native sequence or alternatively, may be  
 5 silent with regard to function.

The *s-ship* promoter sequence of the invention is exemplified by the nucleic acid sequence given in SEQ ID NO:1. Alternatively, an *s-ship* promoter sequence can include all or part of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4 and/or SEQ ID NO:5, as well as any of SEQ ID NO:6-11, or any sequence with at least 80% identity to such sequences and/or  
 10 capable of hybridizing to the complements of such sequences under conditions of high stringency. The invention is not limited to SEQ ID NO:1 as a person of ordinary skill in the art could readily manipulate SEQ ID NO:1 or use all or part of it in subsequent assays and experiments. In certain embodiments, the present invention concerns nucleic acid sequences capable of hybridizing all or parts of SEQ ID NO:1. Parts of SEQ ID NO:1 include, in  
 15 specific embodiments, a *s-ship1* promoter with one or more of the following regions of SEQ ID NO:1: from nucleotide (nt) 49485 to 60914 (11.5 kb-GFP construct); 49485 to 57072, which is 7588 nt (7.6 kb construct); from nt 49485 to 55810, which is 6326 nt (6.3 kb construct); from 49485 to 54755, but lacking 57050 to 57883 (6.2kb-GFP construct); from nt 51389 to 55810, which is 4421 nt (4.4 kb construct); from nt 52199 to 56423, which is 4224  
 20 nt (4.2 kb construct); from nt 53820 to 55810, which is 1990 nt (1.9 kb construct); from nt 54755 to 55810, which is 1055 nt (0.96 kb construct); or from nt 55668 to 55810, which is 142 nt (44 nt construct). It is specifically contemplated that nucleic acids of the invention include those capable of hybridizing to such regions or to subsets of such regions. Moreover, it is contemplated that those nucleic acids capable of hybridizing to such regions may be at  
 25 least 80, 85, 90, 95, 96, 97, 98, 99% or more complementary to all or part of these regions of SEQ ID NOs:1-25.

SEQ ID NO:1 is one sequence for the *ship1* gene. The structure of the gene based on SEQ ID NO:1 is as follows: exon 1 (1-300); exon 2 (4914-4977); exon 3 (44875-45025); exon 4 (47380-47551); exon 5 (49130-49271); exon 6 (55771-55858); exon 7 (69077-  
 30 61057); exon 8 (61175-61246); exon 9 (63231-63354); exon 10 (71113-71219); exon 11 (74821-74923); exon 12 (76837-77033); exon 13 (77601-77718); exon 14 (78653-78749); exon 15 (79268-79403); exon 16 (80787-80894); exon 17 (81041-81129); exon 18 (82789-82870); exon 19 (85604-85693); exon 20 (87766-87879); exon 21 (89288-89370); exon 22 (90735-90822); exon 23 (91701-91850); exon 24 (92863-92959); exon 25 (94708-94983);

exon 26 (97360-97953); and exon 27 (98991-100141). The respective introns lie between the exon sequences. In certain embodiments, the *s-ship* promoter comprises the region spanning intron 5 to intron 6 (inclusive) (referred to as "intron 5/6 region") or sequences from that region. This region is in SEQ ID NO:5. It will be understood that there may be minor sequence differences between different isolates and clones. In such cases, a person of skill the art would recognize corresponding regions between different isolates, clones, and strains.

As used herein, "hybridization", "hybridizes" or "capable of hybridizing" is understood to mean the forming of a double or triple stranded molecule or a molecule with partial double or triple stranded nature. The term "anneal" as used herein is synonymous with "hybridize." The term "hybridization", "hybridize(s)" or "capable of hybridizing" encompasses the terms "stringent condition(s)" or "high stringency" and the terms "low stringency" or "low stringency condition(s)."

As used herein "stringent condition(s)" or "high stringency" are those conditions that allow hybridization between or within one or more nucleic acid strand(s) containing complementary sequence(s), but precludes hybridization of random sequences. Stringent conditions tolerate little, if any, mismatch between a nucleic acid and a target strand. Such conditions are well known to those of ordinary skill in the art, and are preferred for applications requiring high selectivity. Non-limiting applications include isolating a nucleic acid, such as a gene or a nucleic acid segment thereof, or detecting at least one specific mRNA transcript or a nucleic acid segment thereof, and the like.

Stringent conditions may comprise low salt and/or high temperature conditions, such as provided by about 0.02 M to about 0.5 M NaCl at temperatures of about 42°C to about 70°C. It is understood that the temperature and ionic strength of a desired stringency are determined in part by the length of the particular nucleic acid(s), the length and nucleobase content of the target sequence(s), the charge composition of the nucleic acid(s), and to the presence or concentration of formamide, tetramethylammonium chloride or other solvent(s) in a hybridization mixture. It is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide. For example, under highly stringent conditions, hybridization to filter-bound DNA may be carried out in 0.5 M NaHPO<sub>4</sub>, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1 x SSC/0.1% SDS at 68°C (Ausubel *et al.*, 1989).

Conditions may be rendered less stringent by increasing salt concentration and/or decreasing temperature. For example, a medium stringency condition could be provided by



about 0.1 to 0.25M NaCl at temperatures of about 37°C to about 55°C, while a low stringency condition could be provided by about 0.15M to about 0.9M salt, at temperatures ranging from about 20°C to about 55°C. Under low stringent conditions, such as moderately stringent conditions the washing may be carried out for example in 0.2 x SSC/0.1% SDS at 42°C (Ausubel *et al.*, 1989). Hybridization conditions can be readily manipulated depending on the desired results.

It is also understood that these ranges, compositions and conditions for hybridization are mentioned by way of non-limiting examples only, and that the desired stringency for a particular hybridization reaction is often determined empirically by comparison to one or more positive or negative controls. Depending on the application envisioned it is preferred to employ varying conditions of hybridization to achieve varying degrees of selectivity of a nucleic acid towards a target sequence. In a non-limiting example, identification or isolation of a related target nucleic acid that does not hybridize to a nucleic acid under stringent conditions may be achieved by hybridization at low temperature and/or high ionic strength. Such conditions are termed "low stringency" or "low stringency conditions", and non-limiting examples of low stringency include hybridization performed at about 0.15 M to about 0.9 M NaCl at a temperature range of about 20°C to about 50°C. Of course, it is within the skill of one in the art to further modify the low or high stringency conditions to suite a particular application.

In other embodiments, hybridization may be achieved under conditions of, for example, 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM MgCl<sub>2</sub>, 1.0 mM dithiothreitol, at temperatures between approximately 20°C to about 37°C. Other hybridization conditions utilized could include approximately 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, at temperatures ranging from approximately 40°C to about 72°C.

However, in addition to the unmodified *s-ship* promoter sequence of SEQ ID NO:1, the current invention includes derivatives of this sequence and compositions made therefrom. In particular, the present disclosure provides the teaching for one of skill in the art to make and use derivatives of the *s-ship* promoter. For example, the disclosure provides the teaching for one of skill in the art to delimit the functional elements within the *s-ship* promoter and to delete any non-essential elements. Functional elements also could be modified to increase the utility of the sequences of the invention for any particular application. For example, a functional region within the *s-ship* promoter of the invention could be modified to cause or

increase tissue-specific expression. Such changes could be made by site-specific mutagenesis techniques, for example, as described below.

One efficient means for preparing such derivatives comprises introducing mutations into the sequences of the invention, for example, the sequence given in SEQ ID NO:1. Such mutants may potentially have enhanced or altered function relative to the native sequence or alternatively, may be silent with regard to function. It will be understood generally that any embodiment discussed in the application with respect to SEQ ID NO:1 may be applied with respect to any other SEQ ID NO, and vice versa.

Mutagenesis may be carried out at random and the mutagenized sequences screened for function in a trial-by-error procedure. Alternatively, particular sequences that provide the *s-ship* promoter with desirable expression characteristics could be identified and these or similar sequences introduced into other related or non-related sequences via mutation. Similarly, non-essential elements may be deleted without significantly altering the function of the elements. It further is contemplated that one could mutagenize these sequences in order to enhance their utility in expressing transgenes in a particular cell type, for example, in a particular stem cell.

The means for mutagenizing a DNA segment containing an *s-ship* promoter sequence of the current invention are well-known to those of skill in the art. Mutagenesis may be performed in accordance with any of the techniques known in the art, such as, but not limited to, synthesizing an oligonucleotide having one or more mutations within the sequence of a particular regulatory region. In particular, site-specific mutagenesis is a technique useful in the preparation of promoter mutants, through specific mutagenesis of the underlying DNA. The technique further provides a ready ability to prepare and test sequence variants, for example, incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into the DNA. Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Typically, a primer of about 17 to about 75 nucleotides or more in length is preferred, with about 10 to about 25 or more residues on both sides of the junction of the sequence being altered.

In general, the technique of site-specific mutagenesis is well known in the art, as exemplified by various publications. As will be appreciated, the technique typically employs a phage vector which exists in both a single stranded and double stranded form. Typical

vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage are readily commercially available and their use is generally well known to those skilled in the art. Double stranded plasmids also are routinely employed in site directed mutagenesis which eliminates the step of transferring the gene of interest from a plasmid to a  
5 phage.

Site-directed mutagenesis in accordance herewith typically is performed by first obtaining a single-stranded vector or melting apart of two strands of a double-stranded vector which includes within its sequence a DNA sequence that includes the *s-ship* promoter. An oligonucleotide primer bearing the desired mutated sequence is prepared, generally  
10 synthetically. This primer is then annealed with the single-stranded vector, and subjected to DNA polymerizing enzymes such as the *E. coli* polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original non-mutated sequence and the second strand bears the desired mutation. This heteroduplex vector is then used to transform or transfect  
15 appropriate cells, such as *E. coli* cells, and cells are selected which include recombinant vectors bearing the mutated sequence arrangement. Vector DNA can then be isolated from these cells and used for plant transformation. A genetic selection scheme was devised by Kunkel *et al.* (1987) to enrich for clones incorporating mutagenic oligonucleotides. Alternatively, the use of PCR<sup>TM</sup> with commercially available thermostable enzymes such as  
20 *Taq* polymerase may be used to incorporate a mutagenic oligonucleotide primer into an amplified DNA fragment that can then be cloned into an appropriate cloning or expression vector. The PCR<sup>TM</sup>-mediated mutagenesis procedures of Tomic *et al.* (1990) and Upender *et al.* (1995) provide two examples of such protocols. A PCR<sup>TM</sup> employing a thermostable ligase in addition to a thermostable polymerase also may be used to incorporate a  
25 phosphorylated mutagenic oligonucleotide into an amplified DNA fragment that may then be cloned into an appropriate cloning or expression vector.

The preparation of sequence variants of the selected promoter DNA segments using site-directed mutagenesis is provided as a means of producing potentially useful promoter sequences and is not meant to be limiting as there are other ways in which sequence variants  
30 of DNA sequences may be obtained. For example, recombinant vectors encoding the desired promoter sequence may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants.

As used herein, the term "oligonucleotide-directed mutagenesis procedure" refers to template-dependent processes and vector-mediated propagation which result in an increase in the concentration of a specific nucleic acid molecule relative to its initial concentration, or in an increase in the concentration of a detectable signal, such as amplification. As used herein, the term "oligonucleotide directed mutagenesis procedure" also is intended to refer to a process that involves the template-dependent extension of a primer molecule. The term template-dependent process refers to nucleic acid synthesis of an RNA or a DNA molecule wherein the sequence of the newly synthesized strand of nucleic acid is dictated by the well-known rules of complementary base pairing (see, for example, Watson and Ramstad, 1987). Typically, vector mediated methodologies involve the introduction of the nucleic acid fragment into a DNA or RNA vector, the clonal amplification of the vector, and the recovery of the amplified nucleic acid fragment. Examples of such methodologies are provided by U.S. Patent No. 4,237,224, specifically incorporated herein by reference in its entirety. A number of template dependent processes are available to amplify the target sequences of interest present in a sample, such methods being well known in the art and specifically disclosed herein below.

One efficient, targeted means for preparing mutagenized promoters or enhancers relies upon the identification of putative regulatory elements within the target sequence. This can be initiated by comparison with, for example, promoter sequences known to be expressed in a similar manner. Sequences which are shared among elements with similar functions or expression patterns are likely candidates for the binding of transcription factors and are thus likely elements which confer expression patterns. Confirmation of these putative regulatory elements can be achieved by deletion analysis of each putative regulatory region followed by functional analysis of each deletion construct by assay of a reporter gene which is functionally attached to each construct. As such, once a starting promoter or intron sequence is provided, any of a number of different functional deletion mutants of the starting sequence could be readily prepared.

As indicated above, deletion mutants of the *s-ship* promoter also could be randomly prepared and then assayed. With this strategy, a series of constructs are prepared, each containing a different portion of the clone (a subclone), and these constructs are then screened for activity. A suitable means for screening for activity is to attach a deleted promoter construct to a selectable or screenable marker, and to isolate only those cells expressing the marker protein. In this way, a number of different, deleted promoter constructs are identified which still retain the desired, or even enhanced, activity. The

smallest segment which is required for activity is thereby identified through comparison of the selected constructs. This segment may then be used for the construction of vectors for the expression of exogenous protein.

## 1. Vectors

5 Promoter sequences of the invention may be comprised in a vector. The term "vector" is used to refer to a carrier nucleic acid molecule into which a nucleic acid sequence can be inserted for introduction into a cell where it can be replicated. A nucleic acid sequence can be "exogenous," which means that it is foreign to the cell into which the vector is being introduced or that the sequence is homologous to a sequence in the cell but in a  
10 position within the host cell nucleic acid in which the sequence is ordinarily not found. Vectors include plasmids, cosmids, viruses (bacteriophage, animal viruses, and plant viruses), and artificial chromosomes (e.g., YACs). One of skill in the art would be well equipped to construct a vector through standard recombinant techniques, which are described in Sambrook *et al.*, (1989) and Ausubel *et al.*, 1996, both incorporated herein by reference. In  
15 addition to encoding a polypeptide, a vector may encode other polypeptide sequences such as a tag or targetting molecule. Useful vectors encoding such fusion proteins include pIN vectors (Inouye *et al.*, 1985), vectors encoding a stretch of histidines, and pGEX vectors, for use in generating glutathione S-transferase (GST) soluble fusion proteins for later purification and separation or cleavage. A targetting molecule is one that directs the modified polypeptide to a  
20 particular organ, tissue, cell, or other location in a subject's body.

The term "expression vector" refers to a vector containing a nucleic acid sequence coding for at least part of a gene product capable of being transcribed. In some cases, RNA molecules are then translated into a protein, polypeptide, or peptide. In other cases, these sequences are not translated, for example, in the production of antisense molecules, siRNA  
25 molecules or miRNA molecules. In addition to *s-ship* promoter regions, expression vectors can contain a variety of other "control sequences," which refer to nucleic acid sequences necessary for the transcription and possibly translation of an operably linked coding sequence in a particular host organism. In addition to control sequences that govern transcription and translation, vectors and expression vectors may contain nucleic acid sequences that serve  
30 other functions as well and are described *infra*.

In certain embodiments of the invention, the expression vector comprises a virus or engineered vector derived from a viral genome. The ability of certain viruses to enter cells via receptor-mediated endocytosis, to integrate into host cell genome and express viral genes

stably and efficiently have made them attractive candidates for the transfer of foreign genes into mammalian cells (Ridgeway, 1988; Nicolas and Rubenstein, 1988; Baichwal and Sugden, 1986; Temin, 1986). The first viruses used as gene vectors were DNA viruses including the papovaviruses (simian virus 40, bovine papilloma virus, and polyoma) (Ridgeway, 1988; Baichwal and Sugden, 1986) and adenoviruses (Ridgeway, 1988; Baichwal and Sugden, 1986). These have a relatively low capacity for foreign DNA sequences and have a restricted host spectrum. Furthermore, their oncogenic potential and cytopathic effects in permissive cells raise safety concerns. They can accommodate only up to 8 kb of foreign genetic material but can be readily introduced in a variety of cell lines and laboratory animals (Nicolas and Rubenstein, 1988; Temin, 1986).

The retroviruses are a group of single-stranded RNA viruses characterized by an ability to convert their RNA to double-stranded DNA in infected cells; they can also be used as vectors. Other viral vectors may be employed as expression constructs in the present invention. Vectors derived from viruses such as vaccinia virus (Ridgeway, 1988; Baichwal and Sugden, 1986; Coupar *et al.*, 1988) adeno-associated virus (AAV) (Ridgeway, 1988; Baichwal and Sugden, 1986; Hermonat and Muzycska, 1984) and herpesviruses may be employed. They offer several attractive features for various mammalian cells (Friedmann, 1989; Ridgeway, 1988; Baichwal and Sugden, 1986; Coupar *et al.*, 1988; Horwich *et al.*, 1990).

#### a. Promoters and Enhancers

A promoter may be one naturally associated with a gene or sequence, as may be obtained by isolating the 5' non-coding sequences located upstream of the coding segment and/or exon. Such a promoter can be referred to as "endogenous." Similarly, an enhancer may be one naturally associated with a nucleic acid sequence, located either downstream or upstream of that sequence. Alternatively, certain advantages will be gained by positioning the coding nucleic acid segment under the control of a recombinant or heterologous promoter, which refers to a promoter that is not normally associated with a nucleic acid sequence in its natural environment. A recombinant or heterologous enhancer refers also to an enhancer not normally associated with a nucleic acid sequence in its natural environment. Such promoters or enhancers may include promoters or enhancers of other genes, and promoters or enhancers isolated from any other prokaryotic, viral, or eukaryotic cell, and promoters or enhancers not "naturally occurring," *i.e.*, containing different elements of different transcriptional regulatory regions, and/or mutations that alter expression. In addition to producing nucleic

acid sequences of promoters and enhancers synthetically, sequences may be produced using recombinant cloning and/or nucleic acid amplification technology, including PCR™, in connection with the compositions disclosed herein (see U.S. Patent 4,683,202, U.S. Patent 5,928,906, each incorporated herein by reference). Furthermore, it is contemplated the control sequences that direct transcription and/or expression of sequences within non-nuclear organelles such as mitochondria, chloroplasts, and the like, can be employed as well.

Naturally, it may be important to employ a promoter and/or enhancer that effectively directs the expression of the DNA segment in the cell type, organelle, and organism chosen for expression. Those of skill in the art of molecular biology generally know the use of promoters, enhancers, and cell type combinations for protein expression, for example, see Sambrook *et al.* (1989), incorporated herein by reference. The promoters employed may be constitutive, tissue-specific, inducible, and/or useful under the appropriate conditions to direct high level expression of the introduced DNA segment, such as is advantageous in the large-scale production of recombinant proteins and/or peptides. The promoter may be heterologous or endogenous.

In addition to the *s-ship* promoter, other elements/promoters may be employed, in the context of the present invention, to regulate the expression of a gene. Table 1 is a list of other promoters and enhancers that may be used in conjunction with the *s-ship* promoter of the invention; this list also identifies references that indicate how promoters can be evaluated. It is not intended to be exhaustive of all the possible elements involved in the promotion of expression but, merely, to be exemplary thereof. Table 2 provides examples of inducible elements, which are regions of a nucleic acid sequence that can be activated in response to a specific stimulus.

TABLE 1	
Promoter and/or Enhancer	
Promoter/Enhancer	References
Immunoglobulin Heavy Chain	Banerji <i>et al.</i> , 1983; Gilles <i>et al.</i> , 1983; Grosschedl <i>et al.</i> , 1985; Atchinson <i>et al.</i> , 1986, 1987; Imler <i>et al.</i> , 1987; Weinberger <i>et al.</i> , 1984; Kiledjian <i>et al.</i> , 1988; Porton <i>et al.</i> ; 1990
Immunoglobulin Light Chain	Queen <i>et al.</i> , 1983; Picard <i>et al.</i> , 1984
T-Cell Receptor	Luria <i>et al.</i> , 1987; Winoto <i>et al.</i> , 1989; Redondo <i>et al.</i> ; 1990
HLA DQ a and/or DQ $\beta$	Sullivan <i>et al.</i> , 1987

TABLE 1	
Promoter and/or Enhancer	
Promoter/Enhancer	References
$\beta$ -Interferon	Goodbourn <i>et al.</i> , 1986; Fujita <i>et al.</i> , 1987; Goodbourn <i>et al.</i> , 1988
Interleukin-2	Greene <i>et al.</i> , 1989
Interleukin-2 Receptor	Greene <i>et al.</i> , 1989; Lin <i>et al.</i> , 1990
MHC Class II 5	Koch <i>et al.</i> , 1989
MHC Class II HLA-DRA	Sherman <i>et al.</i> , 1989
$\beta$ -Actin	Kawamoto <i>et al.</i> , 1988; Ng <i>et al.</i> , 1989
Muscle Creatine Kinase (MCK)	Jaynes <i>et al.</i> , 1988; Horlick <i>et al.</i> , 1989; Johnson <i>et al.</i> , 1989
Prealbumin (Transthyretin)	Costa <i>et al.</i> , 1988
Elastase I	Omitz <i>et al.</i> , 1987
Metallothionein (MTII)	Karin <i>et al.</i> , 1987; Culotta <i>et al.</i> , 1989
Collagenase	Pinkert <i>et al.</i> , 1987; Angel <i>et al.</i> , 1987
Albumin	Pinkert <i>et al.</i> , 1987; Tronche <i>et al.</i> , 1989, 1990
$\alpha$ -Fetoprotein	Godbout <i>et al.</i> , 1988; Campere <i>et al.</i> , 1989
t-Globin	Bodine <i>et al.</i> , 1987; Perez-Stable <i>et al.</i> , 1990
$\beta$ -Globin	Trudel <i>et al.</i> , 1987
c-fos	Cohen <i>et al.</i> , 1987
c-HA-ras	Triesman, 1986; Deschamps <i>et al.</i> , 1985
Insulin	Edlund <i>et al.</i> , 1985
Neural Cell Adhesion Molecule (NCAM)	Hirsh <i>et al.</i> , 1990
$\alpha_1$ -Antitrypsin	Latimer <i>et al.</i> , 1990
H2B (TH2B) Histone	Hwang <i>et al.</i> , 1990
Mouse and/or Type I Collagen	Ripe <i>et al.</i> , 1989
Glucose-Regulated Proteins (GRP94 and GRP78)	Chang <i>et al.</i> , 1989
Rat Growth Hormone	Larsen <i>et al.</i> , 1986
Human Serum Amyloid A (SAA)	Edbrooke <i>et al.</i> , 1989
Troponin I (TN I)	Yutzey <i>et al.</i> , 1989



TABLE 1	
Promoter and/or Enhancer	
Promoter/Enhancer	References
Platelet-Derived Growth Factor (PDGF)	Pech <i>et al.</i> , 1989
Duchenne Muscular Dystrophy	Klamut <i>et al.</i> , 1990
SV40	Banerji <i>et al.</i> , 1981; Moreau <i>et al.</i> , 1981; Sleight <i>et al.</i> , 1985; Firak <i>et al.</i> , 1986; Herr <i>et al.</i> , 1986; Imbra <i>et al.</i> , 1986; Kadesch <i>et al.</i> , 1986; Wang <i>et al.</i> , 1986; Ondek <i>et al.</i> , 1987; Kuhl <i>et al.</i> , 1987; Schaffner <i>et al.</i> , 1988
Polyoma	Swartzendruber <i>et al.</i> , 1975; Vasseur <i>et al.</i> , 1980; Katinka <i>et al.</i> , 1980, 1981; Tyndell <i>et al.</i> , 1981; Dandolo <i>et al.</i> , 1983; de Villiers <i>et al.</i> , 1984; Hen <i>et al.</i> , 1986; Satake <i>et al.</i> , 1988; Campbell <i>et al.</i> , 1988
Retroviruses	Kriegler <i>et al.</i> , 1982, 1983; Levinson <i>et al.</i> , 1982; Kriegler <i>et al.</i> , 1983, 1984a, b, 1988; Bosze <i>et al.</i> , 1986; Miksicek <i>et al.</i> , 1986; Celander <i>et al.</i> , 1987; Thiesen <i>et al.</i> , 1988; Celander <i>et al.</i> , 1988; Chol <i>et al.</i> , 1988; Reisman <i>et al.</i> , 1989
Papilloma Virus	Campo <i>et al.</i> , 1983; Lusky <i>et al.</i> , 1983; Spandidos and Wilkie, 1983; Spalholz <i>et al.</i> , 1985; Lusky <i>et al.</i> , 1986; Cripe <i>et al.</i> , 1987; Gloss <i>et al.</i> , 1987; Hirochika <i>et al.</i> , 1987; Stephens <i>et al.</i> , 1987
Hepatitis B Virus	Bulla <i>et al.</i> , 1986; Jameel <i>et al.</i> , 1986; Shaul <i>et al.</i> , 1987; Spandau <i>et al.</i> , 1988; Vannice <i>et al.</i> , 1988
Human Immunodeficiency Virus	Muesing <i>et al.</i> , 1987; Hauber <i>et al.</i> , 1988; Jakobovits <i>et al.</i> , 1988; Feng <i>et al.</i> , 1988; Takebe <i>et al.</i> , 1988; Rosen <i>et al.</i> , 1988; Berkhout <i>et al.</i> , 1989; Laspia <i>et al.</i> , 1989; Sharp <i>et al.</i> , 1989; Braddock <i>et al.</i> , 1989
Cytomegalovirus (CMV)	Weber <i>et al.</i> , 1984; Boshart <i>et al.</i> , 1985; Foecking <i>et al.</i> , 1986
Gibbon Ape Leukemia Virus	Holbrook <i>et al.</i> , 1987; Quinn <i>et al.</i> , 1989

TABLE 2		
Inducible Elements		
Element	Inducer	References
MT II	Phorbol Ester (TFA) Heavy metals	Palmiter <i>et al.</i> , 1982; Haslinger <i>et al.</i> , 1985; Searle <i>et al.</i> , 1985; Stuart <i>et al.</i> , 1985; Imagawa <i>et al.</i> , 1987, Karin <i>et al.</i> , 1987; Angel <i>et al.</i> , 1987b; McNeall <i>et al.</i> , 1989
MMTV (mouse mammary tumor virus)	Glucocorticoids	Huang <i>et al.</i> , 1981; Lee <i>et al.</i> , 1981; Majors <i>et al.</i> , 1983; Chandler <i>et al.</i> , 1983; Lee <i>et al.</i> , 1984; Ponta <i>et al.</i> , 1985; Sakai <i>et al.</i> , 1988
$\beta$ -Interferon	poly(rI)x poly(rc)	Tavernier <i>et al.</i> , 1983
Adenovirus 5 E2	E1A	Imperiale <i>et al.</i> , 1984
Collagenase	Phorbol Ester (TPA)	Angel <i>et al.</i> , 1987a
Stromelysin	Phorbol Ester (TPA)	Angel <i>et al.</i> , 1987b
SV40	Phorbol Ester (TPA)	Angel <i>et al.</i> , 1987b
Murine MX Gene	Interferon, Newcastle Disease Virus	Hug <i>et al.</i> , 1988
GRP78 Gene	A23187	Resendez <i>et al.</i> , 1988
$\alpha$ -2-Macroglobulin	IL-6	Kunz <i>et al.</i> , 1989
Vimentin	Serum	Rittling <i>et al.</i> , 1989
MHC Class I Gene H-2kb	Interferon	Blonar <i>et al.</i> , 1989
HSP70	E1A, SV40 Large T Antigen	Taylor <i>et al.</i> , 1989, 1990a, 1990b
Proliferin	Phorbol Ester-TPA	Mordacq <i>et al.</i> , 1989
Tumor Necrosis Factor	PMA	Hensel <i>et al.</i> , 1989
Thyroid Stimulating Hormone $\alpha$ Gene	Thyroid Hormone	Chatterjee <i>et al.</i> , 1989

The identity of tissue-specific promoters or elements, as well as assays to characterize their activity, is well known to those of skill in the art. Examples of such regions include the human LIMK2 gene (Nomoto *et al.* 1999), the somatostatin receptor 2 gene (Kraus *et al.*,

1998), murine epididymal retinoic acid-binding gene (Lareyre *et al.*, 1999), human CD4 (Zhao-Emonet *et al.*, 1998), mouse alpha2 (XI) collagen (Tsumaki, *et al.*, 1998), D1A dopamine receptor gene (Lee, *et al.*, 1997), insulin-like growth factor II (Wu *et al.*, 1997), human platelet endothelial cell adhesion molecule-1 (Almendo *et al.*, 1996), and the SM22 $\alpha$  promoter.

#### b. Initiation Signals and Internal Ribosome Binding Sites

A specific initiation signal also may be required for efficient translation of coding sequences. These signals include the ATG initiation codon or adjacent sequences. Exogenous translational control signals, including the ATG initiation codon, may need to be provided. One of ordinary skill in the art would readily be capable of determining this and providing the necessary signals. It is well known that the initiation codon must be "in-frame" with the reading frame of the desired coding sequence to ensure translation of the entire insert. The exogenous translational control signals and initiation codons can be either natural or synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements.

In certain embodiments of the invention, the use of internal ribosome entry sites (IRES) elements are used to create multigene, or polycistronic, messages. IRES elements are able to bypass the ribosome scanning model of 5'-methylated Cap dependent translation and begin translation at internal sites (Pelletier and Sonenberg, 1988). IRES elements from two members of the picornavirus family (polio and encephalomyocarditis) have been described (Pelletier and Sonenberg, 1988), as well an IRES from a mammalian message (Macejak and Sarnow, 1991). IRES elements can be linked to heterologous open reading frames. Multiple open reading frames can be transcribed together, each separated by an IRES, creating polycistronic messages. By virtue of the IRES element, each open reading frame is accessible to ribosomes for efficient translation. Multiple genes can be efficiently expressed using a single promoter/enhancer to transcribe a single message (see U.S. Patent 5,925,565 and 5,935,819, herein incorporated by reference).

#### c. Multiple Cloning Sites

Vectors can include a multiple cloning site (MCS), which is a nucleic acid region that contains multiple restriction enzyme sites, any of which can be used in conjunction with standard recombinant technology to digest the vector. (See Carbonelli *et al.*, 1999, Levenson *et al.*, 1998, and Cocea, 1997, incorporated herein by reference.) "Restriction enzyme

digestion" refers to catalytic cleavage of a nucleic acid molecule with an enzyme that functions only at specific locations in a nucleic acid molecule. Many of these restriction enzymes are commercially available. Use of such enzymes is widely understood by those of skill in the art. Frequently, a vector is linearized or fragmented using a restriction enzyme that cuts within the MCS to enable exogenous sequences to be ligated to the vector. "Ligation" refers to the process of forming phosphodiester bonds between two nucleic acid fragments, which may or may not be contiguous with each other. Techniques involving restriction enzymes and ligation reactions are well known to those of skill in the art of recombinant technology.

#### d. Splicing Sites

Most transcribed eukaryotic RNA molecules will undergo RNA splicing to remove introns from the primary transcripts. Vectors containing genomic eukaryotic sequences may require donor and/or acceptor splicing sites to ensure proper processing of the transcript for protein expression. (See Chandler *et al.*, 1997, incorporated herein by reference.)

#### e. Termination Signals

The vectors or constructs of the present invention will generally comprise at least one termination signal. A "termination signal" or "terminator" is comprised of the DNA sequences involved in specific termination of an RNA transcript by an RNA polymerase. Thus, in certain embodiments a termination signal that ends the production of an RNA transcript is contemplated. A terminator may be necessary *in vivo* to achieve desirable message levels.

In eukaryotic systems, the terminator region may also comprise specific DNA sequences that permit site-specific cleavage of the new transcript so as to expose a polyadenylation site. This signals a specialized endogenous polymerase to add a stretch of about 200 A residues (polyA) to the 3' end of the transcript. RNA molecules modified with this polyA tail appear to more stable and are translated more efficiently. Thus, in other embodiments involving eukaryotes, it is preferred that that terminator comprises a signal for the cleavage of the RNA, and it is more preferred that the terminator signal promotes polyadenylation of the message. The terminator and/or polyadenylation site elements can serve to enhance message levels and/or to minimize read through from the cassette into other sequences.

Terminators contemplated for use in the invention include any known terminator of transcription described herein or known to one of ordinary skill in the art, including but not

limited to, for example, the termination sequences of genes, such as for example the bovine growth hormone terminator or viral termination sequences, such as for example the SV40 terminator. In certain embodiments, the termination signal may be a lack of transcribable or translatable sequence, such as due to a sequence truncation.

5                                   **f. Polyadenylation Signals**

In expression, particularly eukaryotic expression, one will typically include a polyadenylation signal to effect proper polyadenylation of the transcript. The nature of the polyadenylation signal is not believed to be crucial to the successful practice of the invention, and/or any such sequence may be employed. Preferred embodiments include the SV40  
10 polyadenylation signal and/or the bovine growth hormone polyadenylation signal, convenient and/or known to function well in various target cells. Polyadenylation may increase the stability of the transcript or may facilitate cytoplasmic transport.

**g. Origins of Replication**

In order to propagate a vector in a host cell, it may contain one or more origins of  
15 replication sites (often termed "ori"), which is a specific nucleic acid sequence at which replication is initiated. Alternatively an autonomously replicating sequence (ARS) can be employed if the host cell is yeast.

**h. Selectable and Screenable Markers**

In certain embodiments of the invention, cells containing a nucleic acid construct of  
20 the present invention may be identified *in vitro* or *in vivo* by including a marker in the expression vector. Such markers would confer an identifiable change to the cell permitting easy identification of cells containing the expression vector. Generally, a selectable marker is one that confers a property that allows for selection. A positive selectable marker is one in which the presence of the marker allows for its selection, while a negative selectable marker  
25 is one in which its presence prevents its selection. An example of a positive selectable marker is a drug resistance marker.

Usually the inclusion of a drug selection marker aids in the cloning and identification of transformants, for example, genes that confer resistance to neomycin, puromycin, hygromycin, DHFR, GPT, zeocin and histidinol are useful selectable markers.

30 In addition to markers conferring a phenotype that allows for the discrimination of transformants based on the implementation of conditions, other types of markers including screenable markers such as GFP, whose basis is colorimetric analysis, are also contemplated.

Alternatively, screenable enzymes such as herpes simplex virus thymidine kinase (*tk*), chloramphenicol acetyltransferase (CAT), or luciferase may be utilized. One of skill in the art would also know how to employ immunologic markers, possibly in conjunction with FACS analysis. The marker used is not believed to be important, so long as it is capable of being expressed simultaneously with the nucleic acid encoding a gene product. Further examples of selectable and screenable markers are well known to one of skill in the art.

## 2. Heterologous Sequences

It is contemplated that polynucleotides of the invention include an *s-ship* promoter region controlling the expression of a heterologous nucleic acid sequence. The sequence may be a gene, cDNA sequence or an untranslated sequence, such as an siRNA. The invention is not limited to any specific sequence, but in certain embodiments, the heterologous sequence encodes any of the following proteins or RNAs.

Table 3 below provides different classes of proteins, and in some cases, examples of those proteins.

TABLE 3

Protein Genus	Protein Subgenus	Protein Species	Protein Subspecies
1) Toxins	Ribosome Inhibitory Proteins		
		Geloinin	
		Ricin A Chain	
		Pseudomonas Exotoxin	
		Diphtheria Toxin	
		Mitogillin	
		Saporin	
2) Cytokines/Growth Factors	Interleukins	IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, IL-19	
	TNF		
	LT		
	Interferons	IFN $\alpha$ , IFN $\beta$ , IFN $\gamma$	
	Colony Stimulating Factors	GM-CSF, G-CSF, M-CSF, CSF	
	LIF		

	Fibroblast Growth Factors	bFGF, FGF, FGF-1, FGF-2, FGF-3, FGF-4, FGF-8, FGF-9, FGF-10, FGF-18, FGF-20, FGF, 23	
	VEGF		
3) Enzymes	Oxidoreductases		
	Transferases	Transferring one-carbon groups	Methyltransferases
			Carboxyl and carbamoyltransferases
			Amidinotransferases
		Transferring aldehyde or ketone residues	
		Acyltransferases	Acyltransferases
			Aminoacyltransferases
		Glycosyltransferases	Hexosyltransferases
		Transferring alkyl or aryl groups, other than methyl groups	
		Transferring nitrogenous groups	Transaminases
			Oximinotransferases
		Transferring phosphorous-containing groups	Phosphotransferases
			Diphosphotransferases
			Nucleotidyltransferases
		Transferring sulfur-containing groups	Sulfur-transferases
			Sulfotransferases
			CoA-transferases
		Transferring selenium-containing groups	
	Hydrolases	Acting on ester bonds	
		Glycosylases	
		Acting on ether bonds	
		Acting on peptide bonds (peptide hydrolases)	
		Acting on carbon-nitrogen bonds, other than peptide bonds	
		Acting on acid anhydrides	

		Acting on carbon-carbon bonds.	
		Acting on halide bonds	
		Acting on phosphorus-nitrogen bonds.	
		Acting on sulfur-nitrogen bonds	
		Acting on carbon-phosphorus bonds	
		Acting on sulfur-sulfur bonds	
	Lyases	Carbon-carbon lyases.	
		Carbon-oxygen lyases	
		Carbon-nitrogen lyases	
		Carbon-sulfur lyases	
		Carbon-halide lyases	
		Phosphorus-oxygen lyases	
	Isomerases	Racemases and epimerases	
		Cis-trans-isomerases	
		Intramolecular oxidoreductases	
		Intramolecular transferases (mutases)	
		Phosphotransferases (phosphomutases)	
	Ligases	Forming carbon-oxygen bonds	
		Forming carbon-sulfur bonds	
		Forming carbon-nitrogen bonds.	
		Forming carbon-carbon bonds	
		Forming phosphoric ester bonds	

Other examples include but are not limited to the following:



**a. Cytokines**

Another class of compounds that is contemplated to be operatively linked to a therapeutic polypeptide, such as a toxin, includes interleukins and cytokines, such as interleukin 1 (IL-1), IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15,  $\beta$ -interferon,  $\alpha$ -interferon,  $\gamma$ -interferon, angiostatin, thrombospondin, endostatin, METH-1, METH-2, Flk2/Flt3 ligand, GM-CSF, G-CSF, M-CSF, and tumor necrosis factor (TNF).

**c. Growth Factors**

In other embodiments of the present invention, growth factors or ligands can be complexed with the therapeutic agent. Examples include VEGF/VPF, FGF, TGF $\beta$ , ligands that bind to a TIE, tumor-associated fibronectin isoforms, scatter factor, hepatocyte growth factor, fibroblast growth factor, platelet factor (PF4), PDGF, KIT ligand (KL), colony stimulating factors (CSFs), LIF, and TIMP.

**d. Inducers of Cellular Proliferation**

Another group of proteins that may be used in conjunction with modified proteins of the present invention, such as modified gelonin toxin, comprises proteins that induce cellular proliferation. In some embodiments, the toxin is operatively linked to a ribozyme that can inactivate an inducer of cellular proliferation, while in others, the toxin is linked to the inducer itself. Alternatively, a toxin may be attached to an antibody that recognizes an inducer of cell proliferation.

The commonality of all of these proteins is their ability to regulate cellular proliferation. For example, a form of PDGF, the sis oncogene, is a secreted growth factor. Oncogenes rarely arise from genes encoding growth factors, and at the present, sis is the only known naturally-occurring oncogenic growth factor. In one embodiment of the present invention, it is contemplated that anti-sense mRNA directed to a particular inducer of cellular proliferation is used to prevent expression of the inducer of cellular proliferation.

The proteins FMS, ErbA, ErbB and neu are growth factor receptors. Mutations to these receptors result in loss of regulatable function. For example, a point mutation affecting the transmembrane domain of the Neu receptor protein results in the neu oncogene. The erbA oncogene is derived from the intracellular receptor for thyroid hormone. The modified oncogenic ErbA receptor is believed to compete with the endogenous thyroid hormone receptor, causing uncontrolled growth.

The largest class of oncogenes includes the signal transducing proteins (*e.g.*, Src, Abl and Ras). The protein Src is a cytoplasmic protein-tyrosine kinase, and its transformation from proto-oncogene to oncogene in some cases, results via mutations at tyrosine residue 527. In contrast, transformation of GTPase protein ras from proto-oncogene to oncogene, in one example, results from a valine to glycine mutation at amino acid 12 in the sequence, reducing ras GTPase activity.

The proteins Jun, Fos and Myc are proteins that directly exert their effects on nuclear functions as transcription factors.

#### e. Inhibitors of Cellular Proliferation

The tumor suppressors function to inhibit excessive cellular proliferation. The inactivation of these genes destroys their inhibitory activity, resulting in unregulated proliferation. It is contemplated that toxins may be attached to antibodies that recognize mutant tumor suppressors or wild-type tumor suppressors. Alternatively, a toxin may be linked to all or part of the tumor suppressor. The tumor suppressors p53, p16 and C-CAM are described below.

High levels of mutant p53 have been found in many cells transformed by chemical carcinogenesis, ultraviolet radiation, and several viruses. The p53 gene is a frequent target of mutational inactivation in a wide variety of human tumors and is already documented to be the most frequently mutated gene in common human cancers. It is mutated in over 50% of human NSCLC (Hollstein *et al.*, 1991) and in a wide spectrum of other tumors.

The p53 gene encodes a 393-amino acid phosphoprotein that can form complexes with host proteins such as large-T antigen and E1B. The protein is found in normal tissues and cells, but at concentrations which are minute by comparison with transformed cells or tumor tissue

Wild-type p53 is recognized as an important growth regulator in many cell types. Missense mutations are common for the p53 gene and are essential for the transforming ability of the oncogene. A single genetic change prompted by point mutations can create carcinogenic p53. Unlike other oncogenes, however, p53 point mutations are known to occur in at least 30 distinct codons, often creating dominant alleles that produce shifts in cell phenotype without a reduction to homozygosity. Additionally, many of these dominant negative alleles appear to be tolerated in the organism and passed on in the germ line. Various mutant alleles appear to range from minimally dysfunctional to strongly penetrant, dominant negative alleles (Weinberg, 1991).

Another inhibitor of cellular proliferation is p16. The major transitions of the eukaryotic cell cycle are triggered by cyclin-dependent kinases, or CDK's. One CDK, cyclin-dependent kinase 4 (CDK4), regulates progression through the G<sub>1</sub>. The activity of this enzyme may be to phosphorylate Rb at late G<sub>1</sub>. The activity of CDK4 is controlled by an activating subunit, D-type cyclin, and by an inhibitory subunit, the p16<sup>INK4</sup> has been  
 5 biochemically characterized as a protein that specifically binds to and inhibits CDK4, and thus may regulate Rb phosphorylation (Serrano *et al.*, 1993; Serrano *et al.*, 1995). Since the p16<sup>INK4</sup> protein is a CDK4 inhibitor (Serrano, 1993), deletion of this gene may increase the activity of CDK4, resulting in hyperphosphorylation of the Rb protein. p16 also is known to  
 10 regulate the function of CDK6.

p16<sup>INK4</sup> belongs to a newly described class of CDK-inhibitory proteins that also includes p16<sup>B</sup>, p19, p21<sup>WAF1</sup>, and p27<sup>KIP1</sup>. The p16<sup>INK4</sup> gene maps to 9p21, a chromosome region frequently deleted in many tumor types. Homozygous deletions and mutations of the p16<sup>INK4</sup> gene are frequent in human tumor cell lines. This evidence suggests that the p16<sup>INK4</sup>  
 15 gene is a tumor suppressor gene. This interpretation has been challenged, however, by the observation that the frequency of the p16<sup>INK4</sup> gene alterations is much lower in primary uncultured tumors than in cultured cell lines (Caldas *et al.*, 1994; Cheng *et al.*, 1994; Hussussian *et al.*, 1994; Kamb *et al.*, 1994; Kamb *et al.*, 1994; Mori *et al.*, 1994; Okamoto *et al.*, 1994; Nobori *et al.*, 1995; Orlow *et al.*, 1994; Arap *et al.*, 1995). Restoration of wild-type  
 20 p16<sup>INK4</sup> function by transfection with a plasmid expression vector reduced colony formation by some human cancer cell lines (Okamoto, 1994; Arap, 1995).

Other genes that may be employed according to the present invention include Rb, APC, mda-7, fus-1, FHIT, p16, DCC, NF-1, NF-2, WT-1, MEN-I, MEN-II, zac1, p73, VHL, MMAC1 / PTEN, DBCCR-1, FCC, rsk-3, p27, p27/p16 fusions, p21/p27 fusions, anti-  
 25 thrombotic genes (*e.g.*, COX-1, TFPI), PGS, Dp, E2F, *ras*, *myc*, *neu*, *raf*, *erb*, *fms*, *trk*, *ret*, *gsp*, *hst*, *abl*, E1A, p300, genes involved in angiogenesis (*e.g.*, VEGF, FGF, thrombospondin, BAI-1, GDAIF, or their receptors) and MCC.

Other examples are provided in Table 4 below.

#### f. Regulators of Programmed Cell Death

30 Apoptosis, or programmed cell death, is an essential process for normal embryonic development, maintaining homeostasis in adult tissues, and suppressing carcinogenesis (Kerr *et al.*, 1972). The Bcl-2 family of proteins and ICE-like proteases have been demonstrated to be important regulators and effectors of apoptosis in other systems. The Bcl-2 protein, discovered in association with follicular lymphoma, plays a prominent role in controlling

apoptosis and enhancing cell survival in response to diverse apoptotic stimuli (Bakhshi *et al.*, 1985; Cleary and Sklar, 1985; Cleary *et al.*, 1986; Tsujimoto *et al.*, 1985; Tsujimoto and Croce, 1986). The evolutionarily conserved Bcl-2 protein now is recognized to be a member of a family of related proteins, which can be categorized as death agonists or death antagonists.

Apo2 ligand (Apo2L, also called TRAIL) is a member of the tumor necrosis factor (TNF) cytokine family. TRAIL activates rapid apoptosis in many types of cancer cells, yet is not toxic to normal cells. TRAIL mRNA occurs in a wide variety of tissues. Most normal cells appear to be resistant to TRAIL's cytotoxic action, suggesting the existence of mechanisms that can protect against apoptosis induction by TRAIL. The first receptor described for TRAIL, called death receptor 4 (DR4), contains a cytoplasmic "death domain"; DR4 transmits the apoptosis signal carried by TRAIL. Additional receptors have been identified that bind to TRAIL. One receptor, called DR5, contains a cytoplasmic death domain and signals apoptosis much like DR4. The DR4 and DR5 mRNAs are expressed in many normal tissues and tumor cell lines. Recently, decoy receptors such as DcR1 and DcR2 have been identified that prevent TRAIL from inducing apoptosis through DR4 and DR5. These decoy receptors thus represent a novel mechanism for regulating sensitivity to a pro-apoptotic cytokine directly at the cell's surface. The preferential expression of these inhibitory receptors in normal tissues suggests that TRAIL may be useful as an anticancer agent that induces apoptosis in cancer cells while sparing normal cells. (Marsters *et al.* 1999).

Subsequent to its discovery, it was shown that Bcl-2 acts to suppress cell death triggered by a variety of stimuli. Also, it now is apparent that there is a family of Bcl-2 cell death regulatory proteins which share in common structural and sequence homologies. These different family members have been shown to either possess similar functions to Bcl-2 (*e.g.*, Bcl<sub>XL</sub>, Bcl<sub>w</sub>, Bcl<sub>s</sub>, Mcl-1, A1, Bfl-1) or counteract Bcl-2 function and promote cell death (*e.g.*, Bax, Bak, Bik, Bim, Bid, Bad, Harakiri). It is contemplated that any of these polypeptides, including TRAIL, or any other polypeptides that induce or promote of apoptosis, may be operatively linked to a toxin, or that an antibody recognizing any of these polypeptides may also be attached to a toxin.

Granzyme enzymes are also capable of inducing apoptosis. These include Granzyme A and Granzyme B.

Other examples are provided in Table 4 below.

TABLE 4:

<i>Gene</i>	<i>Source</i>	<i>Human Disease</i>	<i>Function</i>
<b>Growth Factors</b>			
<i>HST/KS</i>	Transfection		FGF family member
<i>INT-2</i>	MMTV promoter Insertion		FGF family member
<i>INT1/WNT1</i>	MMTV promoter Insertion		Factor-like
<i>SIS</i>	Simian sarcoma virus		PDGF B
<b>Receptor Tyrosine Kinases</b>			
<i>ERBB/HER</i>	Avian erythroblastosis virus; ALV promoter insertion; amplified human tumors	Amplified, deleted Squamous cell Cancer; glioblastoma	EGF/TGF- $\alpha$ / Amphiregulin/ Hecacellulin receptor
<i>ERBB-2/NEU/HER-2</i>	Transfected from rat Glioblastomas	Amplified breast, Ovarian, gastric cancers	Regulated by NDF/ Heregulin and EGF-Related factors
<i>FMS</i>	SM feline sarcoma virus		CSF-1 receptor
<i>KIT</i>	HZ feline sarcoma virus		MGF/Steel receptor Hematopoiesis
<i>TRK</i>	Transfection from human colon cancer		NGF (nerve growth Factor) receptor
<i>MET</i>	Transfection from human osteosarcoma		Scatter factor/HGF Receptor
<i>RET</i>	Translocations and point mutations	Sporadic thyroid cancer; familial medullary thyroid cancer; multiple endocrine neoplasias 2A and 2B	Orphan receptor Tyr Kinase
<i>ROS</i>	URII avian sarcoma Virus		Orphan receptor Tyr Kinase
<i>PDGF</i> receptor	Translocation	Chronic Myelomonocytic Leukemia	TEL(ETS-like transcription factor)/ PDGF receptor gene Fusion
<i>TGF-<math>\beta</math></i> receptor		Colon carcinoma mismatch mutation target	
<b>NONRECEPTOR TYROSINE KINASES</b>			
<i>ABL</i>	Abelson MuLV	Chronic myelogenous leukemia translocation with BCR	Interact with RB, RNA polymerase, CRK, CBL
<i>FPS/FES</i>	Avian Fujinami SV;GA FeSV		
<i>LCK</i>	MuLV (murine leukemia virus) promoter insertion		Src family; T cell signaling; interacts CD4/CD8 T cells
<i>SRC</i>	Avian Rous sarcoma Virus		Membrane-associated Tyr kinase with signaling function; activated by receptor kinases
<i>YES</i>	Avian Y73 virus		Src family; signaling
<b>SER/THR PROTEIN KINASES</b>			
<i>AKT</i>	AKT8 murine retrovirus		Regulated by PI(3)K?

<i>Gene</i>	<i>Source</i>	<i>Human Disease</i>	<i>Function</i>
<i>MOS</i>	Maloney murine SV		regulate 70-kd S6 k? GVBD; cystostatic factor; MAP kinase kinase
<i>PIM-1</i>	Promoter insertion Mouse		
<i>RAF/MIL</i>	3611 murine SV; MH2 avian SV		Signaling in RAS Pathway
<b>MISCELLANEOUS CELL SURFACE</b>			
<i>APC</i>	Tumor suppressor	Colon cancer	Interacts with catenins
<i>DCC</i>	Tumor suppressor	Colon cancer	CAM domains
<i>E-cadherin</i>	Candidate tumor Suppressor	Breast cancer	Extracellular homotypic binding; intracellular interacts with catenins
<i>PTC/NBCCS</i>	Tumor suppressor and <i>Drosophila</i> homology	Nevoid basal cell cancer syndrome (Gorline syndrome)	12 transmembrane domain; signals through Gli homologue CI to antagonize hedgehog pathway Signaling
<i>TAN-1 Notch homologue</i>	Translocation	T-ALL	
<b>MISCELLANEOUS SIGNALING</b>			
<i>BCL-2</i>	Translocation	B-cell lymphoma	Apoptosis
<i>CBL</i>	Mu Cas NS-1 V		Tyrosine- Phosphorylated RING finger interact Abl
<i>CRK</i>	CT1010 ASV		Adapted SH2/SH3 interact Abl
<i>DPC4</i>	Tumor suppressor	Pancreatic cancer	TGF- $\beta$ -related signaling Pathway
<i>MAS</i>	Transfection and Tumorigenicity		Possible angiotensin Receptor
<i>NCK</i>			Adaptor SH2/SH3
<b>GUANINE NUCLEOTIDE EXCHANGERS AND BINDING PROTEINS</b>			
<i>BCR</i>		Translocated with ABL in CML	Exchanger; protein Kinase
<i>DBL</i>	Transfection		Exchanger
<i>GSP</i>			
<i>NF-1</i>	Hereditary tumor Suppressor	Tumor suppressor neurofibromatosis	RAS GAP
<i>OST</i>	Transfection		Exchanger
Harvey-Kirsten, N-RAS	HaRat SV; Ki RaSV; Balb-MoMuSV; Transfection	Point mutations in many human tumors	Signal cascade
<i>VAV</i>	Transfection		S112/S113; exchanger
<b>NUCLEAR PROTEINS AND TRANSCRIPTION FACTORS</b>			
<i>BRCA1</i>	Heritable suppressor	Mammary cancer/ovarian cancer	Localization unsettled
<i>BRCA2</i>	Heritable suppressor	Mammary cancer	Function unknown
<i>ERBA</i>	Avian erythroblastosis Virus		Thyroid hormone receptor (transcription)
<i>ETS</i>	Avian E26 virus		DNA binding
<i>EVII</i>	MuLV promotor	AML	Transcription factor

<i>Gene</i>	<i>Source</i>	<i>Human Disease</i>	<i>Function</i>
<i>FOS</i>	Insertion FBI/FBR murine osteosarcoma viruses	Glioma	Transcription factor with c-JUN
<i>GLI</i>	Amplified glioma		Zinc finger; cubitus interruptus homologue is in hedgehog signaling pathway; inhibitory link PTC and hedgehog
<i>HMGI/LIM</i>	Translocation t(3:12) t(12:15)	Lipoma	Gene fusions high mobility group HMGI-C (XT-hook) and transcription factor LIM or acidic domain
<i>JUN</i>	ASV-17		Transcription factor AP-1 with FOS
<i>MLL/VHRX + ELI/MEN</i>	Translocation/fusion ELL with MLL Trithorax-like gene	Acute myeloid leukemia	Gene fusion of DNA- binding and methyl transferase MLL with ELI RNA pol II elongation factor
<i>MYB</i>	Avian myeloblastosis Virus	Burkitt's lymphoma	DNA binding
<i>MYC</i>	Avian MC29; Translocation B-cell Lymphomas; promoter Insertion avian leukosis Virus		DNA binding with MAX partner; cyclin regulation; interact RB?; regulate apoptosis?
<i>N-MYC</i> <i>L-MYC</i> <i>REL</i>	Amplified	Neuroblastoma Lung cancer	NF-κB family transcription factor
	Avian Reticuloendotheliosis Virus		
<i>SKI</i>	Avian SKV770 Retrovirus		Transcription factor
<i>VHL</i>	Heritable suppressor	Von Hippel-Landau syndrome	Negative regulator or elongin; transcriptional elongation complex
<i>WT-1</i>		Wilm's tumor	Transcription factor
<b>CELL CYCLE/DNA DAMAGE RESPONSE</b>			
<i>ATM</i>	Hereditary disorder	Ataxia-telangiectasia	Protein/lipid kinase homology; DNA damage response upstream in P53 pathway
<i>BCL-2</i> <i>FACC</i>	Translocation Point mutation	Follicular lymphoma Fanconi's anemia group C (predisposition leukemia	Apoptosis
<i>FHIT</i>	Fragile site 3p14.2	Lung carcinoma	Histidine triad-related diadenosine 5',3'''- P <sup>1</sup> .p <sup>4</sup> tetraphosphate asymmetric hydrolase
<i>hMLI/MutL</i>		HNPCC	Mismatch repair; MutL Homologue
<i>HMSH2/MutS</i>		HNPCC	Mismatch repair; MutS

<i>Gene</i>	<i>Source</i>	<i>Human Disease</i>	<i>Function</i>
<i>HPMS1</i>		HNPCC	Homologue Mismatch repair; MutL
<i>hPMS2</i>		HNPCC	Homologue Mismatch repair; MutL
<i>INK4/MTS1</i>	Adjacent INK-4B at 9p21; CDK complexes	Candidate MTS1 suppressor and MLM melanoma gene	p16 CDK inhibitor
<i>INK4B/MTS2</i>		Candidate suppressor	p15 CDK inhibitor
<i>MDM-2</i>	Amplified	Sarcoma	Negative regulator p53
<i>p53</i>	Association with SV40 T antigen	Mutated >50% human tumors, including hereditary Li-Fraumeni syndrome	Transcription factor; checkpoint control; apoptosis
<i>PRAD1/BCL1</i>	Translocation with Parathyroid hormone or IgG	Parathyroid adenoma; B-CLL	Cyclin D
<i>RB</i>	Hereditary Retinoblastoma; Association with many DNA virus tumor Antigens	Retinoblastoma; osteosarcoma; breast cancer; other sporadic cancers	Interact cyclin/cdk; regulate E2F transcription factor
<i>XPA</i>		xeroderma pigmentosum; skin cancer predisposition	Excision repair; photo- product recognition; zinc finger

As discussed above, other heterologous sequences include those that can be used as a reporter, such as a screenable or selectable marker. Included in this category are colorimetric or fluorescent reporters (GFP,  $\beta$ -gal, etc.), enzymatic reporters (CAT, luciferase, horseradish peroxidase, etc.), or drug resistance reporters.

Furthermore, it is contemplated that encoded sequences may be fusion proteins, fragments or portions of proteins (including peptides), chimeric proteins, as well as non-protein molecules such as functional RNA molecules (tRNA, rRNA, miRNA, siRNA, antisense, ribozyme, etc.). Such sequences are well known to those of skill in the art.

Moreover, in some embodiments the nucleic acid sequence is a cell-surface marker or an antibody. Cell-surface molecules include, but are not limited to, cluster of differentiation antigens (CD), CTLA, CMRF, cellular adhesion molecules (CAM) molecules such as CD4, CD8, CMRF83, CTLA4, etc. In some embodiments of the invention, these molecules are used to identify cells with stem-cell properties. For example, the cell-surface molecule can be used to identify and/or segregate such cells. FACS analysis can be implemented for this purpose.



## B. Host Cells

The invention include host cells transfected, transformed, or infected with a recombinant nucleic acid sequence discussed in this application. Such host cells would be considered recombinant host cells. The mode of transmission of the nucleic acid sequence  
5 into the host cell is of no significant consequence with respect to the invention; therefore, the terms "transfected," "transformed," and "infected: are used interchangeably unless otherwise specified.

As used herein, the terms "cell," "cell line," and "cell culture" may be used interchangeably. All of these terms also include their progeny, which is any and all  
10 subsequent generations. It is understood that all progeny may not be identical due to deliberate or inadvertent mutations. In the context of expressing a heterologous nucleic acid sequence, "host cell" refers to a prokaryotic or eukaryotic cell, and it includes any transformable organisms that is capable of replicating a vector and/or expressing a heterologous gene encoded by a vector. A host cell can, and has been, used as a recipient for  
15 vectors. A host cell may be "transfected" or "transformed," which refers to a process by which exogenous nucleic acid, such as an *s-ship* promoter sequence, is transferred or introduced into the host cell. A transformed cell includes the primary subject cell and its progeny.

Host cells may be derived from prokaryotes or eukaryotes, including yeast cells,  
20 insect cells, and mammalian cells, depending upon whether the desired result is replication of the vector or expression of part or all of the vector-encoded nucleic acid sequences. Numerous cell lines and cultures are available for use as a host cell, and they can be obtained through the American Type Culture Collection (ATCC), which is an organization that serves as an archive for living cultures and genetic materials (World Wide Web at atcc.org). An  
25 appropriate host can be determined by one of skill in the art based on the vector backbone and the desired result. A plasmid or cosmid, for example, can be introduced into a prokaryote host cell for replication of many vectors. Bacterial cells used as host cells for vector replication and/or expression include DH5 $\alpha$ , JM109, and KC8, as well as a number of commercially available bacterial hosts such as SURE<sup>®</sup> Competent Cells and SOLOPACK<sup>™</sup>  
30 Gold Cells (STRATAGENE<sup>®</sup>, La Jolla, CA). Alternatively, bacterial cells such as *E. coli* LE392 could be used as host cells for phage viruses. Appropriate yeast cells include *Saccharomyces cerevisiae*, *Saccharomyces pombe*, and *Pichia pastoris*.

Examples of eukaryotic host cells for replication and/or expression of a vector include HeLa, NIH3T3, Jurkat, 293, Cos, CHO, Saos, and PC12. Stem cell lines and other immature cell lines are specifically contemplated as suitable host cells of the invention. Many host cells from various cell types and organisms are available and would be known to one of skill in the art. Similarly, a viral vector may be used in conjunction with either a eukaryotic or prokaryotic host cell, particularly one that is permissive for replication or expression of the vector.

Some vectors may employ control sequences that allow it to be replicated and/or expressed in both prokaryotic and eukaryotic cells. One of skill in the art would further understand the conditions under which to incubate all of the above described host cells to maintain them and to permit replication of a vector. Also understood and known are techniques and conditions that would allow large-scale production of vectors, as well as production of the nucleic acids encoded by vectors and their cognate polypeptides, proteins, or peptides.

In certain embodiments of the invention, a host cell refers to a cell obtained from a subject that is transfected, transformed, or infected with an *s-ship* promoter region. The promoter region may integrate into the cell's genome or it may exist in the cell extrachromosomally. Moreover, it may be operably linked to a nucleic acid sequence whose expression is controlled by the *s-ship* promoter region. In further embodiments, the nucleic acid sequence is a heterologous sequence, meaning it is not one associated with an *s-ship* promoter in nature. In some cases, the heterologous sequence is a reporter gene. In other cases, it is a therapeutic or diagnostic nucleic acid, meaning that the resulting transcript or protein can be used as a therapeutic or diagnostic with respect to the host cell. It is contemplated that the host cell may be obtained from subject, transfected or infected with the *s-ship* promoter region, which may or may be operably linked to a therapeutic or diagnostic nucleic acid, and returned back to the subject. This *ex vivo* approach can be used in a variety of contexts, including but not limited to, the treatment of cancer or other hyperproliferative diseases or disorders, autoimmune disease, diseases or disorders involving stem cells, diseases or disorders treatable with stem cells, and/or diseases or disorders caused by protein deficiencies.

In certain embodiments, the *s-ship* promoter regions are employed to direct transcription in cells in a temporal or developmentally specific manner. Therefore, it is contemplated that in some embodiments of the invention, a recombinant host cell contains a heterologous nucleic acid sequence under the control of an *s-ship* promoter region and the

heterologous nucleic acid sequence is initially expressed but after the cell differentiates, expression is limited or eliminated. In some embodiments, stem cells are used for *ex vivo* therapy in which stem cells or a subset of progenitor cells are obtained either from a non-recipient donor or from the recipient themselves, introduced with the *s-ship* promoter-heterologous nucleic acid sequence, and then administered to the subject in which therapy is needed. It is contemplated that the introduced cells could potentially provide expression insofar as they did not become differentiated.

The invention has applicability also with respect to tumor cells as host cells for nucleotides containing *s-ship* promoter regions. The idea of tumor stem cells has been around for more than 40 years, however, more recent technology has given further credible support to this concept. Initially, experiments demonstrated that about one in a million tumor cells can initiate a tumor upon transplantation into autologous hosts.

Two models were proposed to account for these findings; 1) all tumor cells are alike and stochastic events determine whether a cell might be a tumor initiating cell, and 2) not all tumor cells are alike, but a very few (about  $1/10^6$  tumor cells) are inherently capable of tumor initiation on transplantation into a suitable host. These models, termed variously: stochastic vs. hierarchy, nurture vs. nature, probabilistic vs. deterministic, have been applied to understanding many stem/developmental systems. Recently, however, these models of tumor development have been tested using the more refined NOD/SCID mouse for transplantation. Breast tumor-specific cell-surface markers (along with the absence of lineage markers (lin-) of differentiated cells) have been used for identification and isolation of the few tumor initiating cells within the breast tumor mass (Al-Hajj M. Wicha MS. Benito-Hernandez A. Morrison SJ. Clarke MF. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A. 100:3983-8, 2003.). In this case, specific populations capable of tumor transplantation were identified, and these tumor cells were all lineage-minus. Further experiments showed that the same tumor population could be isolated from the transplanted animals, and that the transplanted cells had reestablished the same heterogeneity in transplanted tumor as found in the primary population. These data strongly support the hierarchy model (#2, above) in which specific cells are destined for sustaining the tumorigenic capability. The results above were obtained with breast tumors, but studies in both brain and acute myeloid leukemia (AML) have supported the same characteristics of tumor cells able to initiate tumors on transplantation (Lapidot *et al.*, 1994; Singh *et al.*, 2003; Bhatia *et al.*, 1997).

In the case of human AML, hematopoietic stem cell surface markers are well known (CD34+, CD38-, Lin-, Thy1.1+, c-Kitlo) and this "stem cell" fraction contained the tumor-initiating cells. Results from all three tumor systems examined indicate that a tissue stem cell might be the initial target for tumor formation. The properties of the tumor stem cells identified by transplantation suggest that they sustain the tumor mass by self-replication; however, partial differentiation of the tumor stem cell produces a vast majority of tumor cells, which can no longer sustain the tumor on transplantation. Thus, like normal stem cells, tumor stem cells self-replicate and differentiate producing a larger mass of differentiated but non-transplantable tumor cells.

This theory of tumor stem cells as the primary source of tumor growth has important implications for tumor therapy and potential stem cell therapies for correction of tissue-dependent human diseases. The fact that the *s-SHIP* promoter expresses exclusively in stem/progenitors in the embryo and adult suggests *s-SHIP* protein can be used to further study tumor models and evaluate pathways and to drive expression of agents, including therapeutic and diagnostic agents, in tumor cells, particularly those that qualify as tumor stem cells.

### C. Assays of Transgene Expression

Assays may be employed with the instant invention for determination of the relative efficiency of transgene expression. For example, assays may be used to determine the efficacy of deletion mutants of the *s-ship* promoter in directing expression of exogenous proteins. Similarly, one could produce random or site-specific mutants of the *s-ship* promoter of the invention and assay the efficacy of the mutants in the expression of a given transgene. Alternatively, assays could be used to determine the efficacy of the *s-ship* promoter in directing protein expression when used in conjunction with various different enhancers, terminators or other types of elements potentially used in the preparation of transformation constructs.

For mammals, expression assays may comprise a system utilizing cell lines, or alternatively, whole organisms. Additionally, assays of tissue or developmental specific promoters are generally feasible.

The biological sample to be assayed may comprise nucleic acids isolated from the cells of any plant material according to standard methodologies (Sambrook *et al.*, 1989). The nucleic acid may be genomic DNA or fractionated or whole cell RNA. Where RNA is used,

it may be desired to convert the RNA to a complementary DNA. In one embodiment of the invention, the RNA is whole cell RNA; in another, it is poly-A RNA. Normally, the nucleic acid is amplified.

Depending on the format, the specific nucleic acid of interest is identified in the sample directly using amplification or with a second, known nucleic acid following amplification. Next, the identified product is detected. In certain applications, the detection may be performed by visual means (e.g., ethidium bromide staining of a gel). Alternatively, the detection may involve indirect identification of the product *via* chemiluminescence, radioactive scintigraphy of radiolabel or fluorescent label or even *via* a system using electrical or thermal impulse signals (Affymax Technology; Bellus, 1994).

Following detection, one may compare the results seen in a given sample with a statistically significant reference group of non-transformed control cells. Typically, the non-transformed control cells will be of a genetic background similar to the transformed cells. In this way, it is possible to detect differences in the amount or kind of protein detected in various transformed cells.

As indicated, a variety of different assays are contemplated in the screening of cells or animals of the current invention and associated promoters. These techniques may in cases be used to detect for both the presence and expression of the particular genes as well as rearrangements that may have occurred in the gene construct. The techniques include but are not limited to, fluorescent *in situ* hybridization (FISH), direct DNA sequencing, pulsed field gel electrophoresis (PFGE) analysis, Southern or Northern blotting, single-stranded conformation analysis (SSCA), RNase protection assay, allele-specific oligonucleotide (ASO), dot blot analysis, denaturing gradient gel electrophoresis, RFLP and PCR<sup>TM</sup>-SSCP.

#### 1. Quantitation of Gene Expression with Relative Quantitative RT-PCR<sup>TM</sup>

Reverse transcription (RT) of RNA to cDNA followed by relative quantitative PCR<sup>TM</sup> (RT-PCR<sup>TM</sup>) can be used to determine the relative concentrations of specific mRNA species, for example, an mRNA whose expression is controlled by an *s-ship* promoter. By determining that the concentration of a specific mRNA species varies, it can be shown that the gene encoding the specific mRNA species is differentially expressed. In this way, a promoters expression profile can be rapidly identified, as can the efficacy with which the promoter directs transgene expression.

In PCR<sup>TM</sup>, the number of molecules of the amplified target DNA increase by a factor approaching two with every cycle of the reaction until some reagent becomes limiting.

Thereafter, the rate of amplification becomes increasingly diminished until there is no increase in the amplified target between cycles. If a graph is plotted in which the cycle number is on the X axis and the log of the concentration of the amplified target DNA is on the Y axis, a curved line of characteristic shape is formed by connecting the plotted points.

5 Beginning with the first cycle, the slope of the line is positive and constant. This is said to be the linear portion of the curve. After a reagent becomes limiting, the slope of the line begins to decrease and eventually becomes zero. At this point the concentration of the amplified target DNA becomes asymptotic to some fixed value. This is said to be the plateau portion of the curve.

10 The concentration of the target DNA in the linear portion of the PCR™ amplification is directly proportional to the starting concentration of the target before the reaction began. By determining the concentration of the amplified products of the target DNA in PCR™ reactions that have completed the same number of cycles and are in their linear ranges, it is possible to determine the relative concentrations of the specific target sequence in the original  
15 DNA mixture. If the DNA mixtures are cDNAs synthesized from RNAs isolated from different tissues or cells, the relative abundances of the specific mRNA from which the target sequence was derived can be determined for the respective tissues or cells. This direct proportionality between the concentration of the PCR™ products and the relative mRNA abundances is only true in the linear range of the PCR™ reaction.

20 The final concentration of the target DNA in the plateau portion of the curve is determined by the availability of reagents in the reaction mix and is independent of the original concentration of target DNA. Therefore, the first condition that must be met before the relative abundances of a mRNA species can be determined by RT-PCR™ for a collection of RNA populations is that the concentrations of the amplified PCR™ products must be  
25 sampled when the PCR™ reactions are in the linear portion of their curves.

The second condition that must be met for an RT-PCR™ study to successfully determine the relative abundances of a particular mRNA species is that relative concentrations of the amplifiable cDNAs must be normalized to some independent standard. The goal of an RT-PCR™ study is to determine the abundance of a particular mRNA species  
30 relative to the average abundance of all mRNA species in the sample.

Most protocols for competitive PCR™ utilize internal PCR™ standards that are approximately as abundant as the target. These strategies are effective if the products of the PCR™ amplifications are sampled during their linear phases. If the products are sampled when the reactions are approaching the plateau phase, then the less abundant product

becomes relatively over represented. Comparisons of relative abundances made for many different RNA samples, such as is the case when examining RNA samples for differential expression, become distorted in such a way as to make differences in relative abundances of RNAs appear less than they actually are. This is not a significant problem if the internal  
5 standard is much more abundant than the target. If the internal standard is more abundant than the target, then direct linear comparisons can be made between RNA samples.

The above discussion describes theoretical considerations for an RT-PCR™ assay for plant tissue. The problems inherent in plant tissue samples are that they are of variable quantity (making normalization problematic), and that they are of variable quality  
10 (necessitating the co-amplification of a reliable internal control, preferably of larger size than the target). Both of these problems are overcome if the RT-PCR™ is performed as a relative quantitative RT-PCR™ with an internal standard in which the internal standard is an amplifiable cDNA fragment that is larger than the target cDNA fragment and in which the abundance of the mRNA encoding the internal standard is roughly 5-100 fold higher than the  
15 mRNA encoding the target. This assay measures relative abundance, not absolute abundance of the respective mRNA species.

Other studies may be performed using a more conventional relative quantitative RT-PCR™ assay with an external standard protocol. These assays sample the PCR™ products in the linear portion of their amplification curves. The number of PCR™ cycles that are optimal  
20 for sampling must be empirically determined for each target cDNA fragment. In addition, the reverse transcriptase products of each RNA population isolated from the various tissue samples must be carefully normalized for equal concentrations of amplifiable cDNAs. This consideration is very important since the assay measures absolute mRNA abundance. Absolute mRNA abundance can be used as a measure of differential gene expression only in  
25 normalized samples. While empirical determination of the linear range of the amplification curve and normalization of cDNA preparations are tedious and time consuming processes, the resulting RT-PCR™ assays can be superior to those derived from the relative quantitative RT-PCR™ assay with an internal standard.

One reason for this advantage is that without the internal standard/competitor, all of  
30 the reagents can be converted into a single PCR™ product in the linear range of the amplification curve, thus increasing the sensitivity of the assay. Another reason is that with only one PCR™ product, display of the product on an electrophoretic gel or another display method becomes less complex, has less background and is easier to interpret.

## 2. Marker Gene Expression

Marker genes represent an efficient means for assaying the expression of transgenes. Using, for example, a selectable marker gene, one could quantitatively determine the expression levels in the cell using a construct comprising the selectable marker coding region operably linked to the promoter to be assayed, *e.g.*, an *s-ship* promoter. Alternatively, particular cell types could be exposed to a selective agent and the relative resistance provided in these cells quantified, thereby providing an estimate of the tissue specific expression of the promoter.

Screenable markers constitute another efficient means for quantifying the expression of a given transgene. Potentially any screenable marker could be expressed and the marker gene product quantified, thereby providing an estimate of the efficiency with which the promoter directs expression of the transgene. Quantification can readily be carried out using either visual means, or, for example, a photon counting device.

A preferred screenable marker gene assay for use with the current invention include the use of the screenable marker gene  $\beta$ -galactosidase ( $\beta$ -gal), luciferase, or green fluorescent protein (GFP).

## 3. Purification and Assays of Proteins

One means for determining the efficiency with which a particular transgene is expressed is to purify and quantify a polypeptide expressed by the transgene. Protein purification techniques are well known to those of skill in the art. These techniques involve, at one level, the crude fractionation of the cellular milieu to polypeptide and non-polypeptide fractions. Having separated the polypeptide from other proteins, the polypeptide of interest may be further purified using chromatographic and electrophoretic techniques to achieve partial or complete purification (or purification to homogeneity). Analytical methods particularly suited to the preparation of a pure peptide are ion-exchange chromatography, exclusion chromatography; polyacrylamide gel electrophoresis; and isoelectric focusing. A particularly efficient method of purifying peptides is fast protein liquid chromatography or even HPLC.

Various techniques suitable for use in protein purification will be well known to those of skill in the art. These include, for example, precipitation with ammonium sulphate, PEG, antibodies and the like or by heat denaturation, followed by centrifugation; chromatography steps such as ion exchange, gel filtration, reverse phase, hydroxylapatite and affinity chromatography; isoelectric focusing; gel electrophoresis; and combinations of such and



other techniques. As is generally known in the art, it is believed that the order of conducting the various purification steps may be changed, or that certain steps may be omitted, and still result in a suitable method for the preparation of a substantially purified protein or peptide.

5 There is no general requirement that the protein or peptide being assayed always be provided in their most purified state. Indeed, it is contemplated that less substantially purified products will have utility in certain embodiments. Partial purification may be accomplished by using fewer purification steps in combination, or by utilizing different forms of the same general purification scheme. For example, it is appreciated that a cation-exchange column chromatography performed utilizing an HPLC apparatus will generally  
10 result in a greater “-fold” purification than the same technique utilizing a low pressure chromatography system. Methods exhibiting a lower degree of relative purification may have advantages in total recovery of protein product, or in maintaining the activity of an expressed protein.

It is known that the migration of a polypeptide can vary, sometimes significantly, with  
15 different conditions of SDS/PAGE (Capaldi *et al.*, 1977). It will therefore be appreciated that under differing electrophoresis conditions, the apparent molecular weights of purified or partially purified expression products may vary.

High Performance Liquid Chromatography (HPLC) is characterized by a very rapid separation with extraordinary resolution of peaks. This is achieved by the use of very fine  
20 particles and high pressure to maintain an adequate flow rate. Separation can be accomplished in a matter of minutes, or at most an hour. Moreover, only a very small volume of the sample is needed because the particles are so small and close-packed that the void volume is a very small fraction of the bed volume. Also, the concentration of the sample need not be very great because the bands are so narrow that there is very little dilution  
25 of the sample.

Gel chromatography, or molecular sieve chromatography, is a special type of partition chromatography that is based on molecular size. The theory behind gel chromatography is that the column, which is prepared with tiny particles of an inert substance that contain small pores, separates larger molecules from smaller molecules as they pass through or around the  
30 pores, depending on their size. As long as the material of which the particles are made does not adsorb the molecules, the sole factor determining rate of flow is the size. Hence, molecules are eluted from the column in decreasing size, so long as the shape is relatively constant. Gel chromatography is unsurpassed for separating molecules of different size because separation is independent of all other factors such as pH, ionic strength, temperature,

*etc.* There also is virtually no adsorption, less zone spreading and the elution volume is related in a simple matter to molecular weight.

Affinity Chromatography is a chromatographic procedure that relies on the specific affinity between a substance to be isolated and a molecule that it can specifically bind to.

5 This is a receptor-ligand type interaction. The column material is synthesized by covalently coupling one of the binding partners to an insoluble matrix. The column material is then able to specifically adsorb the substance from the solution. Elution occurs by changing the conditions to those in which binding will not occur (alter pH, ionic strength, temperature, *etc.*).

10 A particular type of affinity chromatography useful in the purification of carbohydrate containing compounds is lectin affinity chromatography. Lectins are a class of substances that bind to a variety of polysaccharides and glycoproteins.

The matrix should be a substance that itself does not adsorb molecules to any significant extent and that has a broad range of chemical, physical and thermal stability. The  
15 ligand should be coupled in such a way as to not affect its binding properties. The ligand should also provide relatively tight binding. And it should be possible to elute the substance without destroying the sample or the ligand. One of the most common forms of affinity chromatography is immunoaffinity chromatography. The generation of antibodies that would be suitable for use in accord with the present invention is well known to those of skill in the  
20 art.

#### D. Methods of Gene Transfer

Suitable methods for nucleic acid delivery to effect expression of compositions of the present invention are believed to include virtually any method by which a nucleic acid (*e.g.*, DNA, including viral and nonviral vectors) can be introduced into an organelle, a cell, a  
25 tissue or an organism, as described herein or as would be known to one of ordinary skill in the art. Such methods include, but are not limited to, direct delivery of DNA such as by injection (U.S. Patents 5,994,624, 5,981,274, 5,945,100, 5,780,448, 5,736,524, 5,702,932, 5,656,610, 5,589,466 and 5,580,859, each incorporated herein by reference), including microinjection (Harland and Weintraub, 1985; U.S. Patent 5,789,215, incorporated herein by  
30 reference); by electroporation (U.S. Patent No. 5,384,253, incorporated herein by reference); by calcium phosphate precipitation (Graham and Van Der Eb, 1973; Chen and Okayama, 1987; Rippe *et al.*, 1990); by using DEAE-dextran followed by polyethylene glycol (Gopal, 1985); by direct sonic loading (Fechheimer *et al.*, 1987); by liposome mediated transfection

(Nicolau and Sene, 1982; Fraley *et al.*, 1979; Nicolau *et al.*, 1987; Wong *et al.*, 1980; Kaneda *et al.*, 1989; Kato *et al.*, 1991); by microprojectile bombardment (PCT Application Nos. WO 94/09699 and 95/06128; U.S. Patents 5,610,042; 5,322,783 5,563,055, 5,550,318, 5,538,877 and 5,538,880, and each incorporated herein by reference); by agitation with  
5 silicon carbide fibers (Kaepler *et al.*, 1990; U.S. Patents 5,302,523 and 5,464,765, each incorporated herein by reference); by desiccation/inhibition-mediated DNA uptake (Potrykus *et al.*, 1985). Through the application of techniques such as these, organelle(s), cell(s), tissue(s) or organism(s) may be stably or transiently transformed.

## **E. Transgenic and Knockout Animals**

### **1. Transgenic Animals**

10 It is further contemplated that transgenic animals are part of the present invention. A transgenic animal of the present invention may involve an animal in which an *s-ship* promoter drives the expression of a transgene. The transgene can be expressed temporally or spatially in a manner different than or the same as a non-transgenic animal. The transgene  
15 may also be heterologous with respect to the host cell or organism, such as, for example, the luciferase gene in a mammalian cell. Moreover, it is contemplated that the transgene may be expressed in a different tissue type or in a different amount or at a different time than the endogenously expressed version of the transgene.

In a general aspect, a transgenic animal is produced by the integration of a given  
20 transgene into the genome in a manner that permits the expression of the transgene, or by disrupting the wild-type gene, leading to a knockout of the wild-type gene. Methods for producing transgenic animals are generally described by Wagner and Hoppe (U.S. Patent No. 4,873,191; which is incorporated herein by reference), Brinster *et al.* (1985; which is incorporated herein by reference in its entirety) and in "Manipulating the Mouse Embryo; A  
25 Laboratory Manual" 2nd edition (eds., Hogan, Beddington, Costantini and Long, Cold Spring Harbor Laboratory Press, 1994; which is incorporated herein by reference in its entirety).

U.S. Patent 5,639,457 is also incorporated herein by reference to supplement the present teaching regarding transgenic pig and rabbit production. U.S. Patents 5,175,384;  
30 5,175,385; 5,530,179, 5,625,125, 5,612,486 and 5,565,186 are also each incorporated herein by reference to similarly supplement the present teaching regarding transgenic mouse and rat production. Transgenic animals may be crossed with other transgenic animals or knockout animals to evaluate phenotype based on compound alterations in the genome.

## 2. Knockout Animals or Cells

The generation of an animal model lacking *s*-ship or a particular nucleic acid (encoding an RNA that is translated or not) is contemplated as part of the present invention to understand further stem cell function. This strategy could also be implemented in cell culture as well.

The lack of activity as a result of the knockout may provoke various types of pathophysiological disturbances in a knockout animal or cell. This can be used to characterize the role or function of a particular gene product at a particular time in development or in a particular cell type. Use of the *s*-ship promoter can be used to drive the expression of the knockout gene such that only certain cells, for example stem cells, may be affected. One method of inhibiting the endogenous expression of a particular gene in an animal is to disrupt the gene in germline cells and produce offspring from these cells. This method is generally known as knockout technology. U.S. Patent No. 5,616,491, incorporated herein by reference in its entirety, generally describes the techniques involved in the preparation of knockout mice, and in particular describes mice having a suppressed level of expression of the gene encoding CD28 on T cells, and mice wherein the expression of the gene encoding CD45 is suppressed on B cells. Pfeffer *et al.* (1993) describe mice in which the gene encoding the tumor necrosis factor receptor p55 has been suppressed. The mice showed a decreased response to tumor necrosis factor signaling. Fung-Leung *et al.* (1991a; 1991b) describe knockout mice lacking expression of the gene encoding CD8. These mice were found to have a decreased level of cytotoxic T cell response to various antigens and to certain viral pathogens such as lymphocytic choriomeningitis virus.

The term "knockout" refers to a partial or complete suppression of the expression of at least a portion of a protein encoded by an endogenous DNA sequence in a cell. The term "knockout construct" refers to a nucleic acid sequence that is designed to decrease or suppress expression of a protein encoded by endogenous DNA sequences in a cell. The nucleic acid sequence used as the knockout construct is typically comprised of: (1) DNA from some portion of the gene (exon sequence, intron sequence, and/or promoter sequence) to be suppressed, in conjunction with all or part of the *s*-ship promoter; and (2) a marker sequence used to detect the presence of the knockout construct in the cell. The knockout construct is inserted into a cell, and integrates with the genomic DNA of the cell in such a position so as to prevent or interrupt transcription of the native DNA sequence. Such insertion usually occurs by homologous recombination (*i.e.*, regions of the knockout construct that are homologous to endogenous DNA sequences hybridize to each other when

the knockout construct is inserted into the cell and recombine so that the knockout construct is incorporated into the corresponding position of the endogenous DNA).

The knockout construct nucleic acid sequence may comprise 1) a full or partial sequence of one or more exons and/or introns of the gene to be suppressed, 2) a full or partial promoter sequence of the gene to be suppressed, or 3) combinations thereof. Typically, the knockout construct is inserted into an embryonic stem cell (ES cell) and is integrated into the ES cell genomic DNA, usually by the process of homologous recombination. This ES cell is then injected into, and integrates with, the developing embryo.

The phenotype of a mouse heterozygous for the knockout may lend clues as to the function and importance of that gene or sequence, as well as contribute an understanding about its physiological relevance, particularly with respect to disease states. Animals completely lacking the targeted gene (homozygous null) may provide additional information. Mice lacking the targeted gene may not be viable, which itself is indicative of the importance of that gene. Should such mice be viable (heterozygous or homozygous nulls), they may be crossed with other transgenic or knockout mice. Furthermore, knock-out mice having any phenotype that resembles a disease state may be used to screen or test therapeutic drugs that slow, modify, or cure conditions. As is known to the skilled artisan, a conditional knockout, wherein the gene is disrupted under certain conditions, is frequently used.

### 3. Conditional Transgenic and Knockdown Animals and Cells

The present invention further contemplates conditional transgenic or knockdown animals (or cells in culture), such as those produced using recombination methods. Bacteriophage P1 Cre recombinase and flp recombinase from yeast plasmids are two non-limiting examples of site-specific DNA recombinase enzymes which cleave DNA at specific target sites (lox P sites for cre recombinase and ftr sites for flp recombinase) and catalyze a ligation of this DNA to a second cleaved site. A large number of suitable alternative site-specific recombinases have been described, and their genes can be used in accordance with the method of the present invention. Such recombinases include the Int recombinase of bacteriophage  $\lambda$  (with or without Xis) (Weisberg *et al.*, 1983), herein incorporated by reference); TpnI and the  $\beta$ -lactamase transposons (Mercier *et al.*, 1990); the Tn3 resolvase (Flanagan and Fennevald, 1989); the yeast recombinases (Matsuzaki *et al.*, 1990); the B. subtilis SpoIVC recombinase (Sato *et al.*, 1990); the Flp recombinase (Schwartz and Sadowski, 1989; Parsons *et al.*, 1990; Golic and Lindquist, 1989; Amin *et al.*, 1990); the Hin recombinase (Glasgow *et al.*, 1989); immunoglobulin recombinases (Malynn *et al.*, 1988);

and the *Cin* recombinase (Haffter and Bickle, 1988; Hubner *et al.*, 1989), all herein incorporated by reference. Such systems are discussed (Echols, 1990; de Villartay, 1988; Craig, 1988; Poyart-Salmeron *et al.*, 1989; Hunger-Bertling *et al.*, 1990; and Cregg and Madden, 1989), all herein incorporated by reference.

5           Of particular interest in the present invention is the Cre recombinase. Cre has been purified to homogeneity, and its reaction with the loxP site has been extensively characterized (Abremski and Hess, 1984), herein incorporated by reference). Cre protein has a molecular weight of 35,000 and can be obtained commercially from New England Nuclear/DuPont. The cre gene (which encodes the Cre protein) has been cloned and expressed (Abremski *et al.*,  
10           1983), herein incorporated by reference). The Cre protein mediates recombination between two loxP sequences (Sternberg *et al.*, 1981), which may be present on the same or different DNA molecule. Because the internal spacer sequence of the loxP site is asymmetrical, two loxP sites can exhibit directionality relative to one another (Hoess and Abremski, 1984). Thus, when two sites on the same DNA molecule are in a directly repeated orientation, Cre  
15           will excise the DNA between the sites (Abremski *et al.*, 1983). However, if the sites are inverted with respect to each other, the DNA between them is not excised after recombination but is simply inverted. Thus, a circular DNA molecule having two loxP sites in direct orientation will recombine to produce two smaller circles, whereas circular molecules having two loxP sites in an inverted orientation simply invert the DNA sequences flanked by the  
20           loxP sites. In addition, recombinase action can result in reciprocal exchange of regions distal to the target site when targets are present on separate DNA molecules.

Recombinases have important application for characterizing gene function in knockout models. When the constructs described herein are used to disrupt limulus clotting factor protease-like genes, a fusion transcript can be produced when insertion of the positive  
25           selection marker occurs downstream (3') of the translation initiation site of the limulus clotting factor protease-like gene. The fusion transcript could result in some level of protein expression with unknown consequence. It has been suggested that insertion of a positive selection marker gene can affect the expression of nearby genes. These effects may make it difficult to determine gene function after a knockout event since one could not discern  
30           whether a given phenotype is associated with the inactivation of a gene, or the transcription of nearby genes. Both potential problems are solved by exploiting recombinase activity. When the positive selection marker is flanked by recombinase sites in the same orientation, the addition of the corresponding recombinase will result in the removal of the positive

selection marker. In this way, effects caused by the positive selection marker or expression of fusion transcripts are avoided.

### III. Proteinaceous Compositions

5 In certain embodiments, the present invention concerns novel compositions comprising at least one proteinaceous molecule, such as s-SHIP1, SHIP1, or a modulator of an *s-ship1* promoter. As used herein, a "proteinaceous molecule," "proteinaceous composition," "proteinaceous compound," "proteinaceous chain" or "proteinaceous material" generally refers, but is not limited to, a protein of greater than about 200 amino acids or the full length endogenous sequence translated from a gene; a polypeptide of greater than about 100 amino acids; and/or a peptide of from about 3 to about 100 amino acids. All the "proteinaceous" terms described above may be used interchangeably herein.

In certain embodiments the size of the at least one proteinaceous molecule may comprise, but is not limited to, about 5, about 6, about 7, about 8, about 9, about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19, about 20, about 21, about 22, about 23, about 24, about 25, about 26, about 27, about 28, about 29, about 30, about 31, about 32, about 33, about 34, about 35, about 36, about 37, about 38, about 39, about 40, about 41, about 42, about 43, about 44, about 45, about 46, about 47, about 48, about 49, about 50, about 51, about 52, about 53, about 54, about 55, about 56, about 57, about 58, about 59, about 60, about 61, about 62, about 63, about 64, about 65, about 66, about 67, about 68, about 69, about 70, about 71, about 72, about 73, about 74, about 75, about 76, about 77, about 78, about 79, about 80, about 81, about 82, about 83, about 84, about 85, about 86, about 87, about 88, about 89, about 90, about 91, about 92, about 93, about 94, about 95, about 96, about 97, about 98, about 99, about 100, about 110, about 120, about 130, about 140, about 150, about 160, about 170, about 180, about 190, about 200, about 210, about 220, about 230, about 240, about 250, about 275, about 300, about 325, about 350, about 375, about 400, about 425, about 450, about 475, about 500, about 525, about 550, about 575, about 600, about 625, about 650, about 675, about 700, about 725, about 750, about 775, about 800, about 825, about 850, about 875, about 900, about 925, about 950, about 975, about 1000, about 1100, about 1200, about 1300, about 1400, about 1500, about 1750, about 2000, about 2250, about 2500 or greater amino molecule residues, and any range derivable therein.

As used herein, an "amino molecule" refers to any amino acid, amino acid derivative or amino acid mimic as would be known to one of ordinary skill in the art. In certain

embodiments, the residues of the proteinaceous molecule are sequential, without any non-amino molecule interrupting the sequence of amino molecule residues. In other embodiments, the sequence may comprise one or more non-amino molecule moieties. In particular embodiments, the sequence of residues of the proteinaceous molecule may be interrupted by one or more non-amino molecule moieties.

Accordingly, the term "proteinaceous composition" encompasses amino molecule sequences comprising at least one of the 20 common amino acids in naturally synthesized proteins, or at least one modified or unusual amino acid.

Proteinaceous compositions may be made by any technique known to those of skill in the art, including the expression of proteins, polypeptides or peptides through standard molecular biological techniques, the isolation of proteinaceous compounds from natural sources, or the chemical synthesis of proteinaceous materials. The nucleotide and protein, polypeptide and peptide sequences for various genes have been previously disclosed, and may be found at computerized databases known to those of ordinary skill in the art. One such database is the National Center for Biotechnology Information's Genbank and GenPept databases (<http://www.ncbi.nlm.nih.gov/>). The coding regions for these known genes may be amplified and/or expressed using the techniques disclosed herein or as would be known to those of ordinary skill in the art. Alternatively, various commercial preparations of proteins, polypeptides and peptides are known to those of skill in the art.

In certain embodiments a proteinaceous compound may be purified. Generally, "purified" will refer to a specific or protein, polypeptide, or peptide composition that has been subjected to fractionation to remove various other proteins, polypeptides, or peptides, and which composition substantially retains its activity, as may be assessed, for example, by the protein assays, as would be known to one of ordinary skill in the art for the specific or desired protein, polypeptide or peptide.

It is contemplated that virtually any protein, polypeptide or peptide containing component may be used in the compositions and methods disclosed herein. However, it is preferred that the proteinaceous material is biocompatible.

#### IV. Therapeutic Applications

The invention is widely applicable to a variety of situations where it is desirable to be able to regulate the level of gene expression, such as by turning gene expression "on" and "off", in a rapid, efficient and controlled manner without causing pleiotropic effects or cytotoxicity. The invention may be particularly useful for gene therapy purposes in humans,



in treatments for either genetic or acquired diseases. The general approach of gene therapy involves the introduction of one or more nucleic acid molecules into cells such that one or more gene products encoded by the introduced genetic material are produced in the cells to restore or enhance a functional activity. For reviews on gene therapy approaches Anderson, *et al.* (1992; Miller *et al.* (1992); Friedmann *et al.* (1989); and Cournoyer *et al.* (1990). However, current gene therapy vectors typically utilize constitutive regulatory elements which are responsive to endogenous transcription factors. These vector systems do not allow for the ability to modulate the level of gene expression in a subject. In contrast, the regulatory system of the invention provides this ability.

10 To use the system of the invention for gene therapy purposes, at least one DNA molecule is introduced into cells of a subject in need of gene therapy (e.g., a human subject suffering from a genetic or acquired disease) to modify the cells. The cells are modified to comprise: 1) nucleic acid encoding an inducible regulator of the invention in a form suitable for expression of the inducible regulator in the host cells; and 2) an siRNA (e.g., for  
15 therapeutic purposes) operatively linked to a tissue-specific promoter such as an *s-shipl* promoter. A single DNA molecule encoding components of the regulatory system of the invention can be used, or alternatively, separate DNA molecules encoding each component can be used. The cells of the subject can be modified *ex vivo* and then introduced into the subject or the cells can be directly modified *in vivo* by conventional techniques for  
20 introducing nucleic acid into cells. Thus, the regulatory system of the invention offers the advantage over constitutive regulatory systems of allowing for modulation of the level of gene expression depending upon the requirements of the therapeutic situation.

Genes of particular interest to be knocked down or knocked out in cells of a subject for treatment of genetic or acquired diseases include those encoding a deleterious gene  
25 product, such as an abnormal protein. Examples of non-limiting specific diseases include anemia, blood-related cancers, Parkinson's disease, and diabetes.

The present invention can be applied to develop autologous or allogeneic cell lines for therapeutical purposes. For example, gene therapy applications of particular interest in cell and/or organ transplantation are utilized with the present invention. In exemplary  
30 embodiments, downregulation of transplantation antigens (such as, for example, by downregulation of beta2-microglobulin expression *via* siRNA) allows for transplantation of allogeneic cells while minimizing the risk of rejection by the patient's immune system. The present invention would allow for a switch off of the RNAi in case of adverse effects (e.g. uncontrollable replication of the transplanted cells).

Cells types that can be subjected to the present invention include hematopoietic stem cells, myoblasts, hepatocytes, lymphocytes, airway epithelium, skin epithelium, islets, dopaminergic neurons, keratinocytes, and so forth. For further descriptions of cell types, genes and methods for gene therapy see *e.g.*, Armentano *et al.* (1990); Wolff *et al.* (1990);  
5 Chowdhury *et al.* (1991); Ferry *et al.* (1991); Quantin *et al.* (1992); Dai *et al.* (1992); van Beusechem *et al.* (1992); Rosenfeld *et al.* (1992); Kay *et al.* (1992); Cristiano *et al.* (1993); Hwu *et al.* (1993); and Herz and Gerard (1993).

In particular embodiments of the present invention, there is a method of treating any disease condition amenable to treatment with an *s-ship* promoter. In specific embodiments,  
10 the method comprises preparing a polynucleotide construct having a region encoding a therapeutic or diagnostic (marker) gene that is operably linked to an *s-ship* promoter, wherein the gene encoded by the construct is for the treatment of the disease condition.

#### A. Pharmaceutical Formulations, Delivery, and Treatment Regimens

In an embodiment of the present invention, methods of treatment are contemplated.  
15 An effective amount of the pharmaceutical composition, generally, is defined as that amount sufficient to detectably and repeatedly to ameliorate, reduce, minimize or limit the extent of the disease or its symptoms. More rigorous definitions may apply, including elimination, eradication or cure of disease.

The routes of administration will vary, naturally, with the location and nature of the  
20 lesion, and include, *e.g.*, intradermal, transdermal, parenteral, intravenous, intramuscular, intranasal, subcutaneous, percutaneous, intratracheal, intraperitoneal, intratumoral, perfusion, lavage, direct injection, and oral administration and formulation.

Solutions of the active compounds as free base or pharmacologically acceptable salts may be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose.  
25 Dispersions may also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations contain a preservative to prevent the growth of microorganisms. The pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions (U.S. Patent  
30 5,466,468, specifically incorporated herein by reference in its entirety). In all cases the form must be sterile and must be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (*e.g.*, glycerol,

propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and/or vegetable oils. Proper fluidity may be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. The prevention of the action of microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

For parenteral administration in an aqueous solution, for example, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intramuscular, subcutaneous, intratumoral and intraperitoneal administration. In this connection, sterile aqueous media that can be employed will be known to those of skill in the art in light of the present disclosure. For example, one dosage may be dissolved in 1 ml of isotonic NaCl solution and either added to 1000 ml of hypodermoclysis fluid or injected at the proposed site of infusion, (see for example, "Remington's Pharmaceutical Sciences" 15th Edition, pages 1035-1038 and 1570-1580). Some variation in dosage will necessarily occur depending on the condition of the subject being treated. The person responsible for administration will, in any event, determine the appropriate dose for the individual subject. Moreover, for human administration, preparations should meet sterility, pyrogenicity, general safety and purity standards as required by FDA Office of Biologics standards.

Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized active ingredients into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum-drying and freeze-drying techniques which yield a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

The compositions disclosed herein may be formulated in a neutral or salt form. Pharmaceutically-acceptable salts, include the acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example,

hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like.

5 Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective. The formulations are easily administered in a variety of dosage forms such as injectable solutions, drug release capsules and the like.

As used herein, "carrier" includes any and all solvents, dispersion media, vehicles, 10 coatings, diluents, antibacterial and antifungal agents, isotonic and absorption delaying agents, buffers, carrier solutions, suspensions, colloids, and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients can also be 15 incorporated into the compositions.

The phrase "pharmaceutically-acceptable" or "pharmacologically-acceptable" refers to molecular entities and compositions that do not produce an allergic or similar untoward reaction when administered to a human. The preparation of an aqueous composition that contains a protein as an active ingredient is well understood in the art. Typically, such 20 compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection can also be prepared.

#### **B. Combination Treatments**

The compounds and methods of the present invention may be used in the context of traditional therapies. In order to increase the effectiveness of a treatment with the 25 compositions of the present invention, it may be desirable to combine these compositions with other agents effective in the treatment of those diseases and conditions. For example, the treatment of a cancer may be implemented with therapeutic compounds of the present invention and other anti-cancer therapies, such as anti-cancer agents or surgery. Likewise, the treatment of a vascular disease or condition may involve both the present invention and 30 conventional vascular agents or therapies.

Various combinations may be employed; for example, a host cell of the present invention is "A" and the secondary anti-cancer agent/therapy is "B":

A/B/A B/A/B B/B/A A/A/B A/B/B B/A/A A/B/B/B B/A/B/B

B/B/B/A B/B/A/B A/A/B/B A/B/A/B A/B/B/A B/B/A/A

5 B/A/B/A B/A/A/B A/A/A/B B/A/A/A A/B/A/A A/A/B/A

Administration of the therapeutic expression constructs of the present invention to a patient will follow general protocols for the administration of that particular secondary therapy, taking into account the toxicity, if any, of the treatment. It is expected that the treatment cycles would be repeated as necessary. It also is contemplated that various standard therapies, as well as surgical intervention, may be applied in combination with the described therapy.

## V. EXAMPLES

The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventor to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

### EXAMPLE 1:

#### Materials and Methods

##### Cell growth and transfection conditions

25 NIH3T3 cells, originally obtained from the American Type Culture Collection (ATCC, Rockville, Maryland), were grown in DMEM with 10% fetal bovine serum. The D3 embryonic stem (ES) cell line was obtained from Dr. Tasuku Honjo (Nakano *et al.*, 1994) and grown in high glucose DMEM (GIBCO/Invitrogen Corp., #11965-092) supplemented with 2 mM L-glutamine, 1 mM sodium pyruvate, 0.1 mM nonessential amino acids, 0.15 mM monothioglycerol (Sigma, M7522), and 15% fetal bovine serum (pre-tested for ES cell growth (HyClone Labs, Inc.)). D3 ES cells were routinely grown on a LIF-producing feeder layer of mitomycin C-treated (Nagy *et al.*, 2003) SNL cells, obtained from Phil Soriano

(FHCRC). The SNL cells are G418-resistant. Usually, one passage before flow cytometry, ES cell were transferred to gelatin(Sigma)-coated plates without a feeder layer and with LIF (ESGRO) added to the medium (1000 units/ml).

DNA was transfected into D3 ES cells by electroporation essentially as described by Nagy *et al.*, (2003). ES cells were suspended in PBS ( $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$ - free) at  $1 \times 10^6$  cells/ml and 0.8 ml of the cell suspension placed in a 0.4-cm-wide electrode-gap sterile cuvette (BIO-RAD). Plasmid DNA (20  $\mu\text{g}$ ), linearized by overnight digestion with *Afl* II and Qiagen-purified, was added and mixed. Two pulses (instead of one as recommended) of current were applied to the cells in the cuvette employing settings of 500mF, and 230V on a BIO-RAD Gene-Pulser<sup>TM</sup> with Capacitance Extender. After 5 min on ice, the viscous solution was transferred to a 10-cm culture dish containing mitomycin C-treated SNL cells: After 24 hr, G418 selection was begun using 280  $\mu\text{g}/\text{ml}$  active G418. Cells were passed after 10-14 days onto gelatin-coated plates (no feeder cells) in LIF containing medium with G418. Flow cytometry was performed 3-4 days later.

*Afl* II-linearized plasmid DNA (10  $\mu\text{g}$ ) was introduced into NIH3T3 cells by transfection using Superfect reagent (Qiagen) as recommended by the manufacturer. G418 selection was begun 24 hr after transfection using 400  $\mu\text{g}/\text{ml}$  G418. Cells were passaged twice in G418 before flow cytometry. Regardless of the electroporation into ES cells or transfection into the NIH3T3 cells, abundant G418 resistant colonies were obtained for each cell type.

Two positive control GFP-expression plasmids were used for both NIH3T3 cells and the D3 ES cells to be sure the transfection/electroporation steps were functional and that GFP expression occurred in each experiment. These positive controls also helped set the gates for analyses of GFP -expressing cells. These two plasmids were the pIRES2-GFP empty plasmid (BD Biosciences Clontech) and pIRES2-GFP containing an insert encoding the Capn5 gene. Both plasmids expressed equally well in each cell type, and the empty pIRES2-GFP vector always expressed higher levels of GFP than the one containing the insert.

#### Immunoblotting analysis for SHIP proteins

The techniques for cell extraction, electrophoresis, and immunoblotting have been described previously (Liu *et al.*, 2001). Equal amounts of protein extracts from each cell type were loaded for gel electrophoresis. SHIP proteins were detected using monoclonal antibody P2C6 at a 1:1000 dilution ( Lucas and Rohrschneider, 1999).

### Flow cytometry

Cells were examined for GFP expression on a Caliber II bench-top analyzer. Cytometer setting were established using positive FDC-P1 cells expressing GFP from a retroviral vector and negative cells, not transfected, or transfected with an empty plasmid. At least  $10^4$  cells were analyzed for each plasmid transfected, and two independent transfections were examined. Both transfections gave similar results, and the results of one experiment are shown.

### Construction of promoter-less GFP-expression constructs for analysis of s-SHIP intron-5 promoter activity

A 7.6-kb DNA *Sac* I-*Sac* I fragment from a Lambda 129Sv mouse genomic clone (Wolf *et al.*, 2000, NCBI accession #AF235499, hereby incorporated by reference) was used for initial examination of potential tissue-specific promoter activity. This region contained almost all of intron-5, the 88 bp of exon-6, and 1271 bp extending into intron-6. This 7.6-kb segment was cloned into pBluescript KS (Stratagene), and sub-segments of the region were obtained with the restriction sites shown in FIG. 2. These sub-segments were cloned into a promoter-less GFP-expression construct.

The promoter-less GFP-expression construct was made from the pEGFP-1 plasmid (BD Biosciences Clontech) by modifications of the MCS (multiple cloning site), incorporating additional synthesized cloning sites (EcoRI-AccI(up)-BssHII-NheI-PstI) for insertion of the sub-fragments from the 7.6 kb intron-5 clone. Both AccI and BssHII recognize multiple sequences and the nucleotide sequence in the synthesized DNA corresponded to AccI site at nucleotide 2776 of the 7.6-kb region, and the 5' BssHII site of the pBluescript plasmid, respectively. In addition, prior to incorporation of the extended MCS, the SV40 early and late introns from pCMV $\beta$  were inserted at the 3' end of the MCS between the KpnI and AgeI sites. Two intron cassettes were used: one containing only the splice acceptor site from the long intron, and a second containing both early and late introns. The former was used only for inserts (*e.g.*, the 7.6-kb and 4.2-kb inserts) containing an intact exon 6 with its splice donor site. The two final plasmids each containing the extended MCS and either the late SV40 intron only (pEGFP2-SD3-1), or both SV40 introns (pEGFP2-SD1-2), were sequenced through the inserted intron region and one of each with correct sequence selected for inserting the 7.6-kb clone and sub-regions.

The longest promoter construct contained the entire 7.6-kb putative *s-SHIP* promoter region, and was excised from the pBluescript plasmid with BssHII for insertion into the MCS of the pEGFP-SD3-1 plasmid. The 6.3-kb fragment was obtained with a partial PstI digestion and complete BssHII digestion. The 4.4-kb and 4.2-kb fragments were from derived from  
5 PstI and AccI digestions, respectively. The 1.9-kb segment was obtained from digestion of the 4.4 kb fragment with NheI. The smallest 0.96 kb region was produced by deleting a region of the pBluescript 7.6 kb clone from the SmaI site 960 nucleotides 5' of exon 6, to the FbaI site 22 nucleotides from the 5' end of the 7.6 kb clone. After ligation, the fragment from  
10 the 5' BssHII site to the PstI site was excised. Each fragment was inserted into their respective restriction sites of the extended MCS. Restriction analysis of each purified plasmid confirmed the correct insert in the correct orientation, and all cloning junctions were sequenced to confirm proper ligation. Each plasmid was linearized with AflIII, and Qiagen purified from agarose gels before electroporation or transfection.

#### Construction of the 11.5kb- and 6.2kb-GFP *s-SHIP* promoter transgenes

15 The 11.5kb-GFP transgenic construct was prepared from two separate plasmids containing the two halves of the proposed *s-SHIP* promoter region, plus an 833 nt sequence from a lambda genomic clone, which was inserted between these two halves. The genomic organization of SHIP1 is shown in Wolf *et al.* (2000). The starting genomic clone contained a 4 kb region from the SacI site near the 3' end of the 7.6 kb genomic clone in intron 6,  
20 extending through exon 8 and into intron 8. This SacI-SacI fragment was cloned into the SacI site of pBluescript SK (pBSK). The GFP gene from pEGFP-1 (Invitrogen/Clontech) was excised with NcoI (encompassing the ATG translation start site of GFP) and SspI. This was ligated into the NcoI (the putative *s-SHIP* translation start site in exon 7) and EcoRV sites of the pBSK-4kb clone. Next, the 5' half of the genomic promoter was added in the form of the  
25 SacI-SacI 7.6 kb genomic sub-clone. This was inserted into the one remaining SacI site at the 5' end of the intron 6-exon 7-GFP clone in pBSK. This left a gap of 0.9 kb between the two SacI sites in intron 6 (see Wolf *et al.*, 2000). This region was recovered as a larger BsiWI-EcoRI 2117 nt fragment, whose sequence demonstrated the insertion of 833 nucleotides between two SacI sites. Therefore, this BsiWI-EcoRI fragment was inserted into the same  
30 unique sites of the transgenic construct to produce the finished 11.5kb-GFP transgene in pBSK.

The 6.2kb-GFP transgene-construct was prepared from the 11.5kb-GFP transgene prior to the insertion of the 833 nt at the intron 6 SacI site. This 11.5kb( $\Delta$ 833)-GFP construct



was digested with FbaI and SwaI, removing 5.3 kb from the 5' end of intron 5. Re-ligation removed all but 19 intron 5 nt at the 5' end of the 11.5kb-GFP transgene. Both 11.5kb-GFP and 6.2kb-GFP transgenes, in pBSK, were cut from the plasmid with BssHII and Qiagen purified from an agarose gel for introduction into the mouse genome.

## 5 Production of transgenic mice

Founder transgenic mice were prepared in our Transgenic Mouse Facility by pronuclear injection of fertilized zygotes from (C57Bl/6 female X CBA/J male) F1 mice. Mice, positive for the transgene, were screened by PCR using DNA obtained from tails or toes of young animals. The location of the primer set for PCR is shown in FIG. 3: the upstream primer (a) is within intron 6 (Pro-up2, 5'-TACTCCTCAGCAAGAGTAGCTGG-3')(SEQ ID NO:12), and the downstream primer (b) within the GFP gene (GFP-dn1, 5'-GCTGAACTTGTGGCCGTTTACGT-3')(SEQ ID NO:13) produce a 632 nucleotide (nt) product. These primers were used for detection of both 6.2kb-GFP and 11.5kb-GFP transgenic mice. Positive chimeric mice were bred to C57Bl/6 mice and four founder lines (A, B, C and D) obtained for the 11.5kb-GFP mice. Later analyses demonstrated that founder line B was not positive for GFP expression, even though the primer pair a and b gave a positive 632 nt product. Therefore, line B is not included in further analyses. The other lines were maintained by breeding transgene-positive animals with wild-type C57Bl/6 mice. For some experiments transgene-positive offspring were generated from positive intra-line breeding. Two founder animals were obtained for the 6.2kb-GFP transgene but one was lost.

The transgene copy number in each founder line (except 11.5kb-GFP, line B) was determined by semi-quantitative RT-PCR of transgene expression relative to endogenous *Gab2* expression. Primers for detecting genomic *gab2* are: E4F, 5'-CTTCTATAGCCTTCCCAAGCC-3' (SEQ ID NO:14); E5R, 5'-CTCGTAGGTCTCACAGGAAG-3' (SEQ ID NO:15).

## Analysis of embryos

Preimplantation embryos were harvested at 2.5 and 3.5 dpc from uterine horns of pregnant females [see Nagy *et al.*, (2003) for details of these methods]. The morulae and blastocysts were washed in RPMI 1640 medium (Gibco) containing 10% fetal bovine serum, transferred to PBS ( $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$ ), and GFP-expression or phase images photographed on a Nikon Eclipse TE200 inverted microscope coupled to a Roper Scientific 1kx1k pixel digital camera. Images were captured with MetaMorph software and prepared for publication with

Photoshop (Adobe). High-resolution z-sections of GFP expression within embryos were made with a Leica TCS SP Confocal microscope.

Several blastocysts were plated onto gelatin-coated tissue-culture wells in DME 10% fetal bovine serum, and photographed three days later. During this period, blastocysts hatched  
 5 from the zona pellucida, and attached to the culture plate. The attached mass of trophectoderm cells with the non-adherent ICM was photographed for GFP and phase with a Nikon Eclipse TE200 microscope.

#### RT-PCR analysis of s-SHIP expression in blastocysts

mRNA was isolated from wild-type 3.5 dpc blastocysts, FDC-P1 cells and the D3 ES  
 10 cells using a Dynabeads mRNA DIRECT micro kit (Dyna). Reverse transcription used the Sensiscript kit from Qiagen, and the PCR cycling conditions were as follows: 94°C 1 min, [94°C 15 sec, 68°C 2 min] x 30 cycles, 68°C 5 min, and a 4°C hold. Each reaction used the equivalent of 1.5 ng mRNA, based on the concentration before reverse transcription. Primers pairs were:

15 HPRT-up1, 5'-CCTGCTGGATTACATTAAAGCACTG-3' (SEQ ID NO:16),  
 HPRT-down1 5'-GTCAAGGGCATATCCAACAACAAAC-3' (SEQ ID NO:17);

OCT4-Up1 5'-GGCGTTCTCTTTGGAAAGGTGTTC-3' (SEQ ID NO:18),  
 OCT4-Down1 5'-CTCGAACCACATCCTTCTCT-3' (SEQ ID NO:19);

20 SHIP1/s-SHIP pair #3,  
 SHIP-E8FW, 5'-TTGCTGCACGAGGGCTCAGAATC-3' (SEQ ID NO:20),  
 SSP883RV, 5'-TCCGATTCTCATGCTCTGGCTTG-3' (SEQ ID NO:21);

25 SHIP1/s-SHIP pair #4,  
 SP2109FW, 5'-CAGCCCTGTCTTTGCCACGTTTG-3' (SEQ ID NO:22),  
 SP2637RV, 5'-TCCACTGGATTCATCCCGCTCTG-3' (SEQ ID NO:23);

SHIP1/s-SHIP pair #5,  
 30 newfw, 5'-CTTCCTCTTGCAACAGAGAACCC-3' (SEQ ID NO:24),  
 newrv, 5'-ACTCAACGTCCACTTTGAGATGC-3' (SEQ ID NO:25).

**EXAMPLE 2:****Identification and Characterization of the s-SHIP Promoter**

Potential *s-SHIP* promoter activity was first analyzed in cell lines grown in culture. Several cell lines were tested for s-SHIP vs. SHIP1 protein expression, based on the known and expected expression pattern of the s-SHIP protein (Lioubin *et al.*, 1994; Tu *et al.*, 2001). These results showed the expression of the ~104-kDa s-SHIP only in the ES cells, whereas the 145-kDa SHIP1 product was exclusively expressed in the maturing FD-Fms myeloid cells. Hot SDS-extraction of the ES cells did not change the size of the s-SHIP protein, suggesting that this 104-kDa product is not the result of proteolytic degradation during extraction (Horn *et al.*, 2001). SHIP proteins were not detectable in NIH3T3 fibroblasts, the SNL cells serving as feeder for the ES cell growth, or the 293 human kidney cells. Therefore, NIH3T3 cells and D3 ES cells were selected as negative and positive cells, respectively, for analysis of the potential *s-SHIP* promoter activity.

A 7.6-kb genomic *ship1* region containing the intron-5 region was obtained for initial promoter analysis. The entire 7.6-kb region and sub-fragments thereof were cloned into a promoter-less GFP (enhanced green-fluorescent protein) expression vector (FIG. 1). Promoter activity of the intron-5 region was then assayed in the cells positive for s-SHIP expression (embryonic stem cells, clone D3) vs. cells negative for s-SHIP expression (NIH3T3 cells). The expression of GFP in each cell type, assayed by flow cytometry, was a measure of the promoter activity within each fragment of the 7.6 kb genomic DNA. The results indicated that, whereas, empty vectors alone lacked significant promoter activity in either cell type, vectors containing intron-5 segments exhibited substantial expression in the D3 ES cells but not in the NIH3T3 cells. Segments of intron 5, ranging from 0.96 kb to 7.6 kb were active for GFP expression in the ES cells; however, the shorter segments appeared most active. Two fragments of 1.9 kb and 0.96 kb, immediately upstream of exon 6, each exhibited equally high GFP expression. The shortest insert fragment contained part of exon 6, but only the 44 nucleotides upstream of exon 6, (Tu *et al.*, 2001), and was completely without promoter activity. These results strongly suggest that the intron-5 region of genomic *ship1* contains cell-specific promoter activity, and segments more distal to exon 6 may have negative regulatory activity.

Based on the ES/NIH3T3 cell-transfection experiments, two new constructs with an extended region downstream of the intron-5 genomic area were prepared for *in vivo* analysis

of promoter activity in transgenic mice (FIG. 3A). Transgenic mice were produced for in vivo examination of the putative s-SHIP promoter/enhancer activity, and determining the overall expression pattern of the transgene, and presumably s-SHIP protein. The promoter in the longer of the new constructs (the 11.5kb-GFP transgene) contained the entire intron 5  
5 from the above 7.6-kb genomic fragment, plus all of exon 6, intron 6, and the portion of exon 7 ending at the theoretical ATG start site (Kozak, 1987) for the s-SHIP protein translation. This start site was fused, in frame, to the ATG for the GFP protein. All of intron 6 and part of exon 7 were included in this construct because, 1) the construct might then more closely resemble the endogenous promoter, 2) splicing may be important for efficient expression  
10 (Nott *et al.*, 2004), and 3) positive or negative regulatory elements for expression may also reside within this sequence. The second, shorter, transgenic promoter construct (the 6.2kb-GFP transgene) was similar, but contained only 0.96 kb of intron 5 sequence adjacent to exon 6, and also lacked 833 nucleotides between two SacI sites within intron 6. Thus, if either construct contained promoter activity *in vivo*, transcription would start within intron 5, while  
15 intron 6 would be spliced out and translation of GFP would begin at the first ATG within an appropriate Kozak site.

Transgenic (Tg) mice were then produced in the Hutchinson Center Transgenic Mouse facility and chimera animals screened for each transgene by PCR. Breeding each founder to wild-type C57Bl/6 mice yielded four lines containing the 11.5kb-GFP transgene,  
20 and one line with the 6.2kb-GFP transgene. Of the four founder Tg11.5 kb-GFP mice, one was negative for expression of the transgene (line B), while three were positive and each has exhibited the same expression patterns (lines A, C and D). Copy numbers of genomic transgenes, measured relative to the endogenous *gab2* gene are shown in FIG. 3B. Within the three GFP-expressing 11.5kb-GFP founder mice, empirical results indicate that line C  
25 exhibits the noticeably highest GFP expression levels. Line C mice also exhibit lower birth rates with *in utero* death at 8.5-9.5 days postcoitum (dpc) apparent. The single 6.2kb-GFP founder line harbors the most transgene copies, but no overt defects in the physical appearance of these mice, their birth rate or development have been observed.

Experiments were then conducted with the adult transgenic 11.5kb-GFP mice to  
30 examine transgene expression; however, it was difficult initially to find any GFP expressed in these mice by flow cytometry of blood and stem cell enriched bone marrow. After several negative attempts to find GFP expression, it was reasoned that because ES cell expression was readily detectable in the initial ES cell experiments, the best test for in vivo expression

would be the inner cell mass (ICM) of the blastocyst, from which ES cells can be derived. Therefore, we looked for GFP expression in 3.5-dpc blastocysts derived from mating of Tg males x WT females. Blastocysts derived from one such cross produced 9 GFP-positive embryos indicating that the Tg was homozygous for the transgene. A separate Tg male bred to a WT female produced both positive and negative blastocysts. GFP-positive morulae were also obtained from similar crosses; whereas, blastocysts or morulae from WT parents were negative for GFP.

Blastocysts are composed of 2-3 cell types depending on their developmental stage. The outer trophoctoderm layer of cells surrounds the eccentric inner cell mass (ICM), destined to become the embryo proper, and later stage blastocysts also contain endodermal cells separating the ICM from the blastocoel cavity (Nagy *et al.*, 2003). To obtain a better idea of which cells of the blastocyst express the GFP transgene, transgenic 3.5-dpc blastocysts were allowed to adhere to a culture dish by three days growth in DME 10% FBS. Under these conditions, the zona pellucida is shed, and the outer trophoctoderm cells of the blastocyst form an adherent layer while the ICM remains as an unspread mass, and each is distinguishable morphologically from the other. The results showed that the ICM portion of the blastocyst retained the GFP expression while the adherent trophoctoderm cells were largely GFP-negative.

A more detailed picture of GFP expression throughout the intact early pre-implantation embryos was seen in confocal Z-sections of GFP within transgenic 2.5-dpc morulae and 3.5-dpc blastocysts. All cells of the 16 to 32-cell morula were GFP-positive. Transition of the morula to the early blastocyst is marked by the formation of the blastocoel cavity. A few cells of this early blastocyst structure began to shut-off GFP expression, and the extent of this GFP shut-off was more evident in the late blastocyst. Here, the outer trophoctoderm cells had noticeably lower GFP expression, and the GFP-positive cells were confined to the ICM. Endodermal cells were not readily apparent. In these images, it is helpful to remember that the half-life of the GFP fluorescence is greater than 24 hr (Tech. Borchure, BD Bioscience ClonTech), and therefore cells, which have stopped expressing GFP, will retain some GFP protein and fluorescence for several days. Twenty-four hours separates the morula from the blastocyst stages; therefore, transgene shut-off early during this time would result in lower but not complete lack of GFP fluorescence late in this time span. The 11.5-kb transgene *s-SHIP* promoter contains the information for both cell-specific

positive expression in morula and ICM of the blastocyst, but also cell-specific shut-off in trophectoderm cells.

Preimplantation embryos from the Tg6.2kb-GFP mice were analyzed next. The transgene in these mice contained only the proximal 0.96-kb region upstream of exon 6, which was necessary for GFP expression in the ES cells. It also lacked 833 nucleotides between two SacI sites of the intron-6 region. GFP expression in the 3.5-dpc blastocyst of the 6.2kb-GFP line was analyzed. Both qualitative and quantitative features of GFP expression in the Tg6.2kb-GFP blastocysts differed from those in the Tg11.5kb-GFP mice. First, GFP expression in the Tg6.2kb-GFP blastocysts was noticeably stronger (at least 5-fold) than that in the Tg11.5kb-GFP blastocysts, as measured by exposure times for obtaining equivalent GFP images in the Nikon digital microscope. Second, and more noticeable was the lack of GFP shut-off in the trophectoderm cells of the blastocyst. No clear demarcation in GFP expression was evident between ICM vs. trophectoderm as seen in the Tg11.5kb-GFP blastocysts.

Blastocysts from the Tg6.2kb-GFP mice were also allowed to adhere to culture plates and GFP expression was examined. Adherent blastocysts from Tg11.5kb-GFP mice were examined simultaneously. Adherent Tg6.2kb-GFP blastocyst expressed GFP in both ICM and trophectoderm cells in a, frequently, haphazard pattern. The Tg11.5kb-GFP adherent blastocysts expressed GFP only in the ICM as observed previously. A comparison of all embryos examined revealed that an increased GFP expression was apparent within the adherent Tg6.2kb-GFP blastocysts relative to the adherent Tg11.5kb-GFP blastocysts. These results were consistent with the promoter analyses performed in the ES cells (FIG. 1), and suggested that the lack of GFP shut-off by the 6.2kb-GFP transgene was due to negative regulatory information found in either one or both regions of the 11.5kb-GFP construct missing from the 6.2kb-GFP transgene.

The data from Tu *et al.* (2001) and that presented herein demonstrated exclusive s-SHIP (rather than SHIP1) expression in ES cells, yet, even though ES cells are derived from the ICM of the blastocyst and the intron 5 *s-SHIP* promoter functioned well in the ICM, it was still not certain whether the ICM actually expressed s-SHIP *in vivo*. Consequently, s-SHIP mRNA expression was then analyzed by RT-PCR, compared to that of the universally expressed HPRT, and the ES cell and ICM-specific Oct4 transcription factor. RNA from blastocysts, FDC-P1 myeloid progenitor cells, and D3 ES cells, was positive for HPRT as expected, and only the blastocysts and ES cells were positive for Oct4. Initially, the s-SHIP-

specific primers similar to those described by Tu *et al.* (2001) was used to test for s-SHIP expression; however, poor results were obtained. The forward primer in this set was moved 3'-ward into the region identical to SHIP1 but weak detection was still obtained. s-SHIP was therefore detected by "subtraction" using primers detecting both s-SHIP and SHIP1 products vs. primers detecting only the SHIP1 product. These primers clearly demonstrated the presence of full-length SHIP1 only in the FDC-P1 cells, and s-SHIP in both blastocysts and ES cells. The weak detectability of s-SHIP may be due to poor hybridization of the primers, degradation of the 5' s-SHIP mRNA ends, or possibly an additional shorter transcription product from the *ship1* gene.

Examination of the minimal 0.96-kb promoter proximal to exon 6 by MatInspector indicated at several transcription-factor binding site potentially active in ES cells and the blastocyst ICM. FIG. 4 shows the first 600 nucleotides of this region upstream of exon 6, with potential transcription factor binding sites and motifs for transcriptional regulation marked. A transcription initiator sequence (Butler and Kadonaga, 2002) straddles the 5' end of the 44 nt SSR, suggesting a transcriptional start site. Paired GATA, or Lmo2 binding sites are present, two overlapping p53 and Oct-binding sites, and a single extended FOX-factor binding region are prominent motifs. The Oct-binding motif is present in similar regions of both the murine and human s-SHIP promoter, suggesting such a factor could be important for ES and ICM expression. The POU factor Oct4 is expressed in ES cells and is part of an enhancer for ES cell-specific expression of target genes (Dailey *et al.*, 1994). Therefore the Oct site could be part of a similar ES cell enhancer region.

The transgene expression in preimplantation embryos raises a question about possible progenitor transgene expression in the oocytes or sperm of the adult, which then give rise to the fertilized embryo. The transcription factor, Oct4, is expressed in adult and embryonic germ cells, as well as the blastocyst ICM and in ES cells (Pesce *et al.*, 1998). The possibility that the 11.5kb-GFP transgene could also be germ cell specific is even more likely given the prominent Oct4 binding motif within the 0.96 kb minimal promoter upstream of exon 6 (see FIG. 4). Therefore, ovaries and testes from 7-8 week old adult Tg11.5kb-GFP mice were harvested and frozen sections stained with Alexa 594-labeled phalloidin for visualizing tissue structure through polymerized actin staining, and endogenous GFP expression. The results of this experiment demonstrated that neither the developing sperm of the testis, nor the developing oocytes of the ovarian follicles expressed GFP. Only blood vessels of the testes and ovaries exhibited specific GFP expression. Therefore, unlike the Oct4 transcription

factor, the 11.5kb-GFP transgene is not a maternally activated gene, must be transcriptionally activated sometime after the germ cells leave the ovary/testis, and before the 2.5-dpc-morula stage of development.

### **EXAMPLE 3:**

#### **Further Characterization of the *s-SHIP* Promoter**

5 The transgenic mice that were generated from the experiments described in Example 2 were further analyzed by immunofluorescence. Embryos were harvested, washed, and fixed 2-4 hr in 2% paraformaldehyde in PBS, then washed in 30% sucrose in PBS and stored in that solution overnight at 4° C. Embryos were frozen in O.C.T. on dry ice and stored at -80°  
10 C until sectioned. Twenty mm sections on Superfrost/Plus (Fisherbrand) microscope glass slides were air-dried overnight and stored, desiccated, at -20 C. Sections were routinely stained with a rabbit anti-GFP antibody coupled to Alexa 488 (Molecular Probes) to enhance the transgenic GFP detection. For general screening, sections were also stained with phalloidin coupled to Alexa 594 (molecular Probes) for detecting morphology by filamentous  
15 actin staining. Other antibodies used were specific for: CD45-Cy-Chrome labeled, and Flk1 (VEGFR2) phycoerythrin labeled (both from Pharmingen); Oct4, mouse monoclonal (Santa Cruz Biotechnology); E-cadherin, rat monoclonal (Zymed); and alpha smooth muscle actin, mouse monoclonal (Sigma). The BM alkaline phosphatase detection reagent was from Roche. The tissue sections were blocked in 5% fetal bovine serum for 30 min. washed in  
20 PBS, treated 10 min in 0.5% TX-100 in PBS then washed again in PBS. The primary antibodies were applied and sections incubated at RT in a humidified chamber for 45-60 min. Sections were washed 3-times 10 min each in PBS and secondary antibodies added and incubated as before. Final washing was 3-times 15 min and sections were mounted in ProLong (Molecular Probes). All secondary antibodies were from Molecular probes and  
25 labeled with Alexa 594 or Alexa 633. Slides were viewed with a Leitz TCS SP Confocal microscope.

s-SHIP expression was tested in blastocysts by RT-PCR, compared to that of the universally expressed HPRT, and the ES cell and ICM-specific Oct4 transcription factor (Pesce et al., 1998). FDC-P1 myeloid progenitor cells express SHIP1 but not s-SHIP (Lucas and Rohrschneider, 1999; Tu et al. 2001), while conversely, D3 ES cells express only s-SHIP  
30 (Tu et al. 2001; data not shown). These two cell types represented the positive and negative controls. RNA from FDC-P1, D3 ES cells and 3.5 dpc blastocysts were tested and each was



positive for HPRT, while mRNA from the blastocysts and ES cells, but not from the FDC-P1 cells, was positive for Oct4, as expected. s-SHIP was detected by "subtraction" using primers common to both s-SHIP and SHIP1 vs. primers detecting only the SHIP1 product. These primers demonstrate the presence of full-length SHIP1 only in the FDC-P1 cells, and s-SHIP  
5 in both blastocysts and ES cells. Analysis of transgene expression in the post-implantation mouse embryo.

Ongoing development of the blastocyst following implantation of the embryo continues with the formation of the epiblast (the embryo body) from cells comprising the blastocyst ICM. Consistent with this derivation, the epiblast in the Tg11.5kb-GFP E6 embryo  
10 retained GFP expression; however, within 24 hr, epiblast GFP expression was lost. E7-7.5 embryos exhibited individual GFP-positive cells, or groups of cells, in the extraembryonic membranes, often appearing to trail from the epiblast. By 8.5 dpc, GFP expression could no longer be seen in the embryo body itself, but both the yolk sac and placenta contained numerous GFP-positive cells.

Serial sections through the transgenic E8.5 decidua and growing embryo have not detected specific GFP expression in any tissue or cells of the embryo body. Specifically, GFP-positive migrating primordial germ cells (PGCs) have not been detected in the hindgut region of the E8.5 embryos (not shown, but see later results). Within the extraembryonic regions, however, the ectoplacental plate (also called chorioallantoic plate) contained patches  
20 of GFP-positive cells, and both the blood islands and endoderm cell layer of the yolk sac contained individual or small groups of GFP-positive cells. Occasional intense GFP<sup>+</sup> round cells within the primitive erythrocyte-filled blood islands were observed, and groups of GFP<sup>+</sup> cells adjacent to blood islands were also seen. GFP<sup>+</sup> cells of the peripheral ectoplacental plate sometimes appeared contiguous with GFP<sup>+</sup> endodermal cells of the yolk sac. Maternal  
25 placental contributions did not account for the yolk sac nor ectoplacental plate expression profiles, because the same GFP expression patterns were observed in embryos from Tg males mated to WT females (not shown). This GFP expression pattern in the yolk sac is similar to that reported using an enhancer derived from the Scf(Tal1) stem cell protein (Sánchez et al., 1999), suggesting that these GFP<sup>+</sup> cells may be related to hemangioblasts. However,  
30 Scf(Tal1) expression was not detected in the E8.5 GFP<sup>+</sup> extraembryonic membrane cells.

In contrast to the lack of GFP expression within the 8.5-dpc embryo proper, whole-mount observations of 11.5-dpc Tg11.5kb-GFP embryo showed dramatic GFP expression in

the caudal region of the embryos and a distinct pattern of GFP expression on, and around, each hindlimb/forelimb pair. This complex pattern represented multiple distinct expression sites. At this age of development, the strongest and most broadly observed site of GFP expression was the epidermal cell layer of the developing skin. This was seen in whole  
5 mounts but demonstrated best in frozen sections of E11.5 transgenic embryos. The E11.5 transgenic embryos exhibited extensive GFP expression in the body and limbs, but such expression was not seen in the head region at this embryonic stage. A second epidermis-related GFP expression site was the apical ectodermal ridge (AER) of each developing limb, and a third distinct pattern was the mammary buds. These latter structures form by  
10 invaginations of the skin epidermis and were observed as five GFP-positive spots between each forelimb/hindlimb pair, corresponding to the bilateral three thoracic and two inguinal developing mammary glands. These were seen in whole mount embryos or underneath the dissected epidermis.

An additional prominent GFP expression site in 11.5-dpc Tg11.5kb-GFP embryos  
15 was the genital ridge where PGCs were accumulating. The dissected aorta-gonad-mesonephros (AGM) region from an E13.5 embryo demonstrated the localization of the GFP<sup>+</sup> cells within the gonads. The PGCs co-express the nuclear transcription factor Oct4 within the GFP<sup>+</sup> cells, identifying these GFP<sup>+</sup> cells as PGCs. GFP expression was not detected above background in the liver and dorsal aorta. The dorsal aorta is also present in the dissected AGM  
20 between the two gonads, but specific GFP expression was not observable at this site. The absence of GFP expression in the dorsal aorta suggests a lack of any relationship to definitive hematopoiesis, reportedly arising from this region (Dzierzak, Medvinsky, and de Bruijn, 1998).

#### *Primordial germ cells*

25 GFP expression in the PGCs of the Tg11.5kb-GFP embryo was examined in more detail during E9.5 – E18.5 stages of embryonic development. Using Oct4 as a marker for PGCs, the temporal co-expression of GFP and Oct4 was followed from embryonic day 9.5 to 18.5. Consistent with earlier results, GFP was not expressed in the earliest E9.5 PGCs migrating along the hindgut, although Oct4 was present in their nuclei (Fig. 5Aa,b,c). At  
30 E11.5, PGCs of the genital ridges contained GFP and nuclear Oct4. Both E13.5 and E17.5 PGCs were GFP<sup>+</sup> and Oct4<sup>+</sup>. Together with the earlier analyses of the E8.5 embryos, these results suggest that GFP is not expressed in the early migrating PGCs of the E8.5-9.5 embryo,

but is readily detected in the PGCs of the genital ridge and gonads of the E11.5-17.5 embryos.

GFP expression was observed in PGCs from an E13.5 embryo ovary. Many of the GFP<sup>+</sup> PGCs cells at this stage were in cell division. Primitive seminiferous tubules are distinguishable in the E15.5 testes, populated with GFP<sup>+</sup> PGCs and developing spermatogenic cells (Kaufman, 2001). The brightest GFP<sup>+</sup> cells at this stage were in mitosis and these may represent the type A spermatogonia undergoing cell division (Kaufman, 2001). The seminiferous tubules of the E18.5 embryo were, likewise, filled with GFP<sup>+</sup> spermatogenic cells. Sertoli cells attached to the basement membrane of the seminiferous tubule were GFP-negative. Germ cells in both ovaries and testes were positive for GFP in the E13.5-E18.5 stages of embryo development. However, quite surprisingly, neither ovaries nor testes of adults expressed GFP in any stage of germ cell formation. These observations indicate that the expression of the 11.5kb-GFP transgene is positively regulated during the E11.5-18.5 developmental stages, but negatively regulated in the adult.

*11.5kb-GFP transgene expression in other tissues from E15.5-E18.5 embryos*

Following E11.5 in the Tg11.5kb-GFP embryo, new GFP<sup>+</sup> structures are observed. Specifically, the developing cornea of the 15.5-dpc embryo eye was GFP<sup>+</sup>. The cornea is formed from an outer epithelial layer and an inner cell layer derived from neural crest cells. Only the outer epithelial layer showed GFP expression<sup>+</sup>. The retina, derived from neuroepithelium, did not express GFP, while the lens, although an invagination of the epidermis, was likewise GFP-negative.

E15.5 embryos began to exhibit GFP expression in cells surrounding blood vessels, and this expression became more noticeable at E18.5. The expression was restricted to smaller vessels and not expressed in larger arteries (e.g., aorta) or veins. The vessel-associated GFP was expressed in cells wrapping around the circumference of the vessels. This characteristic suggests the GFP<sup>+</sup> cells are smooth muscle cells, and this notion was supported by the observation that *alpha*-smooth muscle actin co-localized with the GFP vessel cells. GFP was not detectable in all vascular smooth muscle cells (vSMCs), and was also not detected in cardiac or skeletal muscle smooth muscle cells. This indicates a highly specific regulation of the 11.5kb-GFP transgene only in a population of the smooth muscle cells associated with small blood vessels.

A few scattered cells in the E15.5 thymus were also positive for GFP. Morphologically, these cells did not appear to be blood derived, but rather adhered to the thymus stroma via E-cadherin, suggesting an epithelial nature.

The paired vomeronasal (Jacobson's) organ in the nasal septum also expressed GFP.  
5 Jacobson's organ is derived from the neuroepithelium, again suggesting a potential connection between transgene expression and epidermal-derived tissues.

The E17.5-18.5 transgenic embryos exhibited GFP expression in cells associated with the forming bone matrix. These small irregularly shaped GFP<sup>+</sup> cells are attached to the newly formed bone and are most likely the osteoblasts in the process of converting extracellular  
10 matrix material to bone (Komori, et al., 1997; Long, et al., 2004). This notion was confirmed by demonstrating the colocalization of alkaline phosphatase and GFP in these cells. Trabecular bone also exhibited alkaline phosphatase-positive/GFP<sup>+</sup> osteoblasts.

*Transgene expression in embryonic epidermis, tissues derived from epidermis, and other epithelial cells*

15 Embryonic skin is a major stem/progenitor cell population essential for formation of several appendages and tissues, such as, hair follicles, sweat glands, mammary tissue, and (more indirectly) prostate. Therefore, these epidermal-derived tissues were examined for GFP expression in Tg11.5kb-GFP embryonic mice.

Hair follicle formation initiates around E13.5 by reciprocal inductive signals between  
20 skin epidermis and underlying mesenchyme (Miller, 2002). The resultant hair follicle placodes exhibit a localized epidermal thickening in a geometrically ordered array due to inhibitory signaling from each placode (Andl et al., 2002). The results demonstrated that E13.5 transgenic embryos exhibit a geometric array of GFP<sup>+</sup> speckles in the skin, suggesting GFP expression in the hair follicle placodes. Frozen sections of skin at this stage revealed  
25 GFP<sup>+</sup> epidermal thickening indicating placode formation. Growth and extension of the placodes into the mesenchyme produces the hair follicle sheath with a distal bulb (Alonso and Fuchs, 2003). Early stages of hair follicle formation showed the incorporation of GFP<sup>+</sup> skin epidermal cells into the follicle extending about one cell diameter into the mesenchyme. The dermal component of the follicle (the dermal papilla) is GFP<sup>-</sup>, and clearly visible in figure 7C  
30 adjacent to the GFP<sup>+</sup> cells. As the follicle extends further, the GFP<sup>+</sup> cells remain as a small cluster, perhaps maintaining their epidermal-cell niche. During further growth of the hair follicle, the GFP<sup>+</sup> cell cluster remained intact at the distal bulb end of the growing follicle,

and the follicle retained the E-cadherin expression, as did the epidermis from which it was derived. Frequently, the base of each growing hair follicle (anchored in the epidermis) lacked significant GFP expression, and the GFP<sup>+</sup> cluster resides at the distal bulb end of the follicle.

5 Mammary bud formation also occurs from skin epidermis by epidermis-mesenchyme induction, but differs in the time of initial formation, the size of placodes, and morphology of the developing tissue. The mammary buds were visible in whole mount immunofluorescence of the E11.5 transgenic embryos, and frozen sections taken at the same stage show GFP<sup>+</sup> mammary placodes. These placodes became round-up and extended into the mesenchyme forming the buds, which subsequently formed larger bulbous structures as they grew. The results also demonstrated the size of a hair follicle relative to a mammary bud at this stage. 10 This growing embryonic mammary tissue exhibited a few GFP<sup>+</sup> cells at the outside periphery of the tissue. The expanding mammary tissue remained attached to the epidermis and the nipple formed around this attachment site.

A third tissue derived from epithelium is the prostate organ. However, unlike hair 15 follicles and mammary buds, prostate is formed from urothelium by poorly understood signals at the base of the bladder in the E17.5 male mouse. Frozen sections of the prostatic region of the E18.5 Tg11.5kb-GFP male mouse demonstrated specific GFP expression in regions destined to become lateral lobes of the prostate. Serial sections through this region showed GFP expression only in the prostatic region, and non-transgenic mice lacked this expression. Overall, these results indicated that epidermal/epithelial tissue constitutes a major 20 target for expression of the 11.5kb-GFP transgene, and tissues derived from epithelium appear to obtain (and maintain) stem cells from these sources.

*Expression of GFP from the shorter 6.2kb-GFP transgene in the mouse embryo*

The 6.2kb-GFP transgene mice were produced for the purpose of obtaining 25 information about expression capability of promoter segments compared to the 11.5kb-GFP transgene mice. In pre-implantation embryos the shorter transgene was strongly expressed throughout the blastocyst and no delineation was apparent between the ICM and trophectoderm cells in the intact blastocyst, or in blastocysts allowed to adhere to tissue culture plastic. This lack of ICM specificity could be due to the higher copy number in the single Tg6.2kb-GFP founder, and/or to the higher expression of the transgene. Regardless, 30 later developmental stages exhibited highly tissue-specific expression patterns, and dramatic differences in the 6.2kb-GFP transgene expression compared to the 11.5kb-GFP transgene

were observed. The qualitative differences were not yet apparent in E8.5 embryos; however, at this stage the Tg 6.2kb-GFP E8.5 embryos no longer expressed GFP ubiquitously but only cells of the extraembryonic membranes were GFP<sup>+</sup>. This expression was similar to that observed with the same stage Tg11.5kb-GFP embryos; however, yolk sac expression was not observed in the Tg6.2kb-GFP mice. Surprisingly, unlike the Tg11.5kb-GFP mice, the E11.5-13.5 Tg6.2kb-GFP embryos were devoid of significant GFP expression. Thus, at these developmental stages two of the most prominent GFP expression sites in the Tg11.5kbFP embryos (i.e., skin epidermis and PGCs in the genital ridge and gonads) were completely absent in the Tg6.2kb-GFP embryos.

At still later developmental times, E18.5, two of the GFP expression sites seen in the Tg11.5kb-GFP embryos were observed in the Tg6.2kb-GFP embryos. These sites were the blood vessels and the osteoblast cells attached to the forming bone matrix. GFP was expressed in the smaller blood vessels, but not all small vessels expressed GFP.

Therefore, the expression of the 6.2kb-GFP transgene within the embryo was limited to fewer tissues than observed with the Tg11.5kb-GFP mice, and represented a subset of expression sites seen within the Tg11.5kb-GFP mice at the same developmental stage. These results suggest that the promoter sequence remaining within the 6.2kb-GFP transgene contains the information (enhancers) for tissue-specific expression of GFP in cells of the extraembryonic membranes at E8.5 and in the smooth muscle cells surrounding blood vessels, and osteoblast in late stage embryos. Conversely, the genetic sequences present in the 11.5kb-GFP transgene, but absent from the 6.2kb-GFP transgene, contain important instructions (perhaps in conjunction with the shorter promoter sequence) for tissue-specific expression in PGCs, skin epidermal cells as well as tissues derived from skin epidermis.

The results presented here demonstrate that the intron-5/6 region of the *shp1* gene contains the promoter/enhancer for tissue-specific expression in primitive embryonic cell populations. GFP expression by the 11.5kb-GFP construct was first observed in cultured ES cells, then in all blastomeres of the morula, and in the ICM of the blastocyst from Tg11.5kb-GFP mice. Thus the initial embryo expression occurred uniformly in the totipotent cells of the preimplantation embryo. Following implantation, the initially GFP<sup>+</sup> epiblast lost GFP expression; however, at E7.5 a few cells in the extraembryonic membranes and placenta were GFP<sup>+</sup>. Observations at different times indicated these cells originated near the epiblast, and probably gave rise to the few GFP<sup>+</sup> yolk sac cells seen in the blood islands beginning at E8.5.

The embryo proper lacked GFP expression from E7.5 to about E10.5, but strong GFP expression was observable around E11.5 in skin epidermis, mammary buds, developing gonads, limb AER region, and a few days later, the developing hair follicles. From E15.5-18.5 the AER GFP label (and structure) vanished, but skin epidermal cells retained GFP expression. During this time, however, GFP<sup>+</sup> cells of the skin appendages were retained in a small cluster (hair follicles) or a few peripheral epithelial cells (mammary tissue). Also, vSMCs, osteoblasts of developing bone, the vomeronasal organ, prostate, and a few cells of the thymus were GFP<sup>+</sup> at E15.5-18.5.

Several significant points can be made about the embryonic expression pattern of the 11.5kb-GFP transgene. First, a strong preference exists for stem/progenitor cells (ES cells, morula, ICM, primordial germ cells, epidermis), but also for several cell types of yet undefined character and potential (extraembryonic cells, yolk sac cells, thymus, vomeronasal cells, vSMCs, osteoblasts). Second, no clear continuum of GFP-expressing cells is observed throughout embryo development; rather, it is likely that transgene expression is turned on and off at various stages and locations during development. Third, as observed in the Tg11.5kb-GFP vs. the Tg6.2kb-GFP mice, distinct portions of the intron5/6 promoter/enhancer are essential for tissue-specific expression. Finally, transgene GFP expression was never observed in more mature cells of either embryonic or adult tissues, and many of the GFP<sup>+</sup> stem/progenitor cells in the embryo are also retained in the adult tissues (ms in preparation). These results indicate that s-SHIP expression is spatially and temporally regulated throughout development (see supplementary Table 5).

TABLE 5

Summary of temporal and spatial GFP expression in the  
Tg11.5kb-GFP and Tg6.2kb-GFP mouse embryos.

Embryo Age	Expression Site	Tg11.5kb-GFP	Tg6.2kb-GFP
E2.5	morula	+++	++++
E3.5	blastula:		
	ICM	+++	+++
	trophect	+/-	++
E6	epiblast	+++	
	extra embryonic	+/-	
E7.5	epiblast	+/-	
	extra embryonic	++	+++
E8.5	yolk sac	++	
	placenta	++	
	PGC (migrating)	-	
E9.5	PGC (migrating)	-	
E11.5	PGC (genital ridge)	++ +++	- -
	skin epidermis	+++	-
	hair follicles	+++	-
	mammary buds	+++	-
	AER	+++	-
E13.5	gonads	+++	-
	epidermis	+++	-
E15.5	gonads: testis	+++	
	ovary	+++	-
	skin epidermis	+++	-
	mammary tissue	+++	-
	thymus	+	-
	blood vessels (SMC)	+/-	++
E17.5-18.5	gonads: testis	+++	-
	ovary	++	
	skin epidermis	+++	-
	hair follicles	+++	-
	mammary tissue	+++	-
	prostate	+++	
	thymus	+	
	blood vessels (SMC)	+	+++
	osteoblasts	++	+++

5

Although several tissues expressed the GFP transgene, most tissues and organs did not, and these include tissues with defined or postulated stem/progenitor cell activity, such as



muscle, pancreas, small intestine and colon (Marshak et al., 2001; Charge and Rudnicki, 2003; for a contrasting view Dor et al., 2004). Also negative for GFP expression were the E11.5 dorsal aorta and E13.5 fetal liver - sites, respectively, where definitive hematopoiesis is proposed to occur, and where primitive hematopoietic stem cells home and develop  
5 (Dzierzak, Medvinsky, and de Bruijn, 1998). This indicates that the 11.5kb-GFP transgenic GFP promoter/enhancer is highly tissue-specific in its activity.

#### **EXAMPLE 4:**

##### **Isolation and Characterization of the Human *s-SHIP* Promoter**

The human *s-ship* promoter has been isolated and compared to the mouse promoter.  
10 FIG. 6 provides a comparison between a region having the genomic sequence of the human promoter that includes 560 nucleotides upstream from exon 6 (at the 3' end of intron 5) and the corresponding sequence from the mouse sequence.

The mouse and human promoters were significantly homologous. Both promoters contain a binding motif for p53 proteins (p53, p63 or p73). The motif is identified in FIG. 6.  
15 Electrophoretic mobility shift assays show that p53 from nuclear extracts of ES cells will bind to a sequence with the p53 motif shown in FIG. 6. FIG. 7 shows p53 binding sequences in mouse, including the different half sites.

In separate experiments, DNA damage caused by UV or gamma irradiation of ES cells induced both p53 and s-SHIP protein expression. ES cells that lack p53, from knockout  
20 experiments, did not express nor induce s-SHIP protein after UV irradiation. From the Tg11.5kb-GFP mice it is known that GFP expression occurs in cells that express p53 (ES cells and others), p63 (epithelial cells of skin, prostate, and mammary tissue), and p73 (neuroepithelial cells as in the vomeronasal organs).

\*\*\*\*\*

25

All of the compositions and/or methods and/or apparatus disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be  
30 applied to the compositions and/or methods and/or apparatus and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents that are

both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

**REFERENCES**

The following references are specifically incorporated herein by reference.

5

U.S. Patent 4,237,224

U.S. Patent 4,683,202

U.S. Patent 4,873,191

U.S. Patent 5,175,384

10

U.S. Patent 5,175,385

U.S. Patent 5,302,523

U.S. Patent 5,322,783

U.S. Patent 5,384,253

U.S. Patent 5,399,363

15

U.S. Patent 5,464,765

U.S. Patent 5,466,468

U.S. Patent 5,530,179

U.S. Patent 5,538,877

U.S. Patent 5,538,880

20

U.S. Patent 5,543,158

U.S. Patent 5,550,318

U.S. Patent 5,563,055

U.S. Patent 5,565,186

U.S. Patent 5,580,859

25

U.S. Patent 5,589,466

U.S. Patent 5,610,042

U.S. Patent 5,612,486

U.S. Patent 5,616,491

U.S. Patent 5,616,491

30

U.S. Patent 5,625,125

U.S. Patent 5,639,457

U.S. Patent 5,641,515

U.S. Patent 5,656,610

U.S. Patent 5,702,932

- U.S. Patent 5,736,524  
U.S. Patent 5,780,448  
U.S. Patent 5,789,215  
U.S. Patent 5,846,225  
5 U.S. Patent 5,846,233  
U.S. Patent 5,925,565  
U.S. Patent 5,928,906  
U.S. Patent 5,935,819  
U.S. Patent 5,945,100  
10 U.S. Patent 5,981,274  
U.S. Patent 5,994,624
- Abbott *et al.*, *J. Nucl. Cardiol.*, 10(4):403-412, 2003.  
Abremski and Hess, *J. Mol. Biol.*, 259:1509-1514, 1984.  
15 Abremski *et al.*, *Cell*, 32:1301-1311, 1983.  
Alonso and Fuchs, *Cell*, 17:1189-1200, 2003.  
Al-Hajj *et al.*, *Proc. Natl. Acad. Sci. USA*, 100:3983-3988, 2003.  
Almendro *et al.*, *J. Immunol.*, 157(12):5411-5421, 1996.  
Amin *et al.*, *J. Molec. Biol.*, 214:55-72, 1990.  
20 Anderson *et al.*, *J. of Immunotherapy*, 12:19-31, 1992.  
Andl *et al.*, *Cell*, 2:643-653, 2002.  
Angel *et al.*, *Cell*, 49:729, 1987b.  
Angel *et al.*, *Mol. Cell. Biol.*, 7:2256, 1987a.  
Arap *et al.*, *Cancer Res.*, 55(6):1351-1354, 1995.  
25 Armentano *et al.*, *Proc. Natl. Acad. Sci. USA*, 87(16):6141-6145, 1990.  
Atchison and Perry, *Cell*, 46:253, 1986.  
Atchison and Perry, *Cell*, 48:121, 1987.  
Ausubel *et al.*, In: *Current Protocols in Molecular Biology*, John, Wiley & Sons, Inc, NY, .  
1996.  
30 Ausubel *et al.*, In: *Current Protocols in Molecular Biology*, John, Wiley & Sons, Inc, NY,  
1994.  
Ausubel *et al.*, In: *Current Protocols in Molecular Biology*, John, Wiley & Sons, Inc, NY,  
1992.

- Baichwal and Sugden, *In: Gene Transfer*, Kucherlapati (Ed.), NY, Plenum Press, 117-148, 1986.
- Bakhshi *et al.*, *Cell*, 41(3):899-906, 1985.
- Banerji *et al.*, *Cell*, 27(2 Pt 1):299-308, 1981.
- 5 Banerji *et al.*, *Cell*, 33(3):729-740, 1983.
- Bellus, *J. Macromol. Sci. Pure Appl. Chem.*, A31(1): 1355-1376, 1994.
- Bellus, *J. Macromol. Sci. Pure Appl. Chem.*, A31(1): 1355-1376, 1994.
- Berkhout *et al.*, *Cell*, 59:273-282, 1989.
- Bhatia *et al.*, *Proc. Natl. Acad. Sci. USA*, 94:5320-5325, 1997.
- 10 Blonar *et al.*, *EMBO J.*, 8:1139, 1989.
- Bodine and Ley, *EMBO J.*, 6:2997, 1987.
- Boshart *et al.*, *Cell*, 41:521, 1985.
- Bosze *et al.*, *EMBO J.*, 5(7):1615-1623, 1986.
- Braddock *et al.*, *Cell*, 58:269, 1989.
- 15 Brinster *et al.*, *Proc. Natl. Acad. Sci. USA*, 82(13):4438-4442, 1985.
- Bulla and Siddiqui, *J. Virol.*, 62:1437, 1986.
- Butler and Kadonaga, *Genes Dev.*, 16:2583-2592, 2002.
- Caldas *et al.*, *Nat. Genet.*, 8(1):27-32, 1994.
- Campbell and Villarreal, *Mol. Cell. Biol.*, 8:1993, 1988.
- 20 Campere and Tilghman, *Genes and Dev.*, 3:537, 1989.
- Campo *et al.*, *Nature*, 303:77, 1983.
- Capaldi *et al.*, *Biochem. Biophys. Res. Comm.*, 74(2):425-433, 1977.
- Carbonelli *et al.*, *FEMS Microbiol. Lett.*, 177(1):75-82, 1999.
- Celander and Haseltine, *J. Virology*, 61:269, 1987.
- 25 Celander *et al.*, *J. Virology*, 62:1314, 1988.
- Chandler *et al.*, *Cell*, 33:489, 1983.
- Chandler *et al.*, *Proc. Natl. Acad. Sci. USA*, 94(8):3596-601, 1997.
- Chang *et al.*, *Mol. Cell. Biol.*, 9:2153, 1989.
- Charge and Rudnicki, *Cell*, 113:422-423, 2003.
- 30 Chatterjee *et al.*, *Proc. Natl. Acad. Sci. USA*, 86:9114, 1989.
- Chen and Okayama, *Mol. Cell Biol.*, 7(8):2745-2752, 1987.
- Cheng *et al.*, *Cancer Res.*, 54(21):5547-5551, 1994.
- Choi *et al.*, *Cell*, 53:519, 1988.
- Chowdhury *et al.*, *Science*, 254(5039):1802-1805, 1991.

- Cleary and Sklar, *Proc. Natl. Acad. Sci. USA*, 82(21):7439-7443, 1985.
- Cleary *et al.*, *J. Exp. Med.*, 164(1):315-320, 1986.
- Coccea, *Biotechniques*, 23(5):814-816, 1997.
- Cohen *et al.*, *J. Cell. Physiol.*, 5:75, 1987.
- 5 Costa *et al.*, *Mol. Cell. Biol.*, 8:81, 1988.
- Coupar *et al.*, *Gene*, 68:1-10, 1988.
- Cournoyer *et al.*, *Curr. Opin. Biotechnol.*, 1(2):196-208, 1990.
- Craig, *Ann. Rev. Genet.*, 22:77-105, 1988.
- Cregg and Madden, *Mol. Gen. Genet.*, 219:320-323, 1989.
- 10 Cripe *et al.*, *EMBO J.*, 6:3745, 1987.
- Cristiano *et al.*, *Proc. Natl. Acad. Sci. USA*, 90(24):11548-11552, 1993.
- Culotta and Hamer, *Mol. Cell. Biol.*, 9:1376, 1989.
- Dai *et al.*, *J. Biol. Chem.*, 267(27):19565-19571, 1992.
- Dailey *et al.*, *Mol. Cell. Biol.*, 14:7758-7769, 1994.
- 15 Daley *et al.*, *Exp. Hematol.*, 21(6):734-737, 1993.
- Dandolo *et al.*, *J. Virology*, 47:55-64, 1983.
- de Villartay, *Nature*, 335:170-174, 1988.
- De Villiers *et al.*, *Nature*, 312(5991):242-246, 1984.
- Deschamps *et al.*, *Science*, 230:1174-1177, 1985.
- 20 Dor *et al.*, *Nature*, 429:41-46, 2004.
- Dzierzak *et al.*, *Immunol. Today*, 19:228-236, 1998.
- Echols, *J. Biol. Chem.*, 265:14697-14700, 1990.
- Edbrooke *et al.*, *Mol. Cell. Biol.*, 9:1908, 1989.
- Edlund *et al.*, *Science*, 230:912-916, 1985.
- 25 Fechheimer, *et al.*, *Proc Natl. Acad. Sci. USA*, 84:8463-8467, 1987.
- Feng and Holland, *Nature*, 334:6178, 1988.
- Ferry *et al.*, *Proc. Natl. Acad. Sci. USA*, 88(19):8377-8381, 1991.
- Firak and Subramanian, *Mol. Cell. Biol.*, 6:3667, 1986.
- Flanagan and Fennewald, *J. Molec. Biol.*, 206:295-304, 1989.
- 30 Foecking and Hofstetter, *Gene*, 45(1):101-105, 1986.
- Fraley *et al.*, *Proc. Natl. Acad. Sci. USA*, 76:3348-3352, 1979.
- Friedmann, *Science*, 244:1275-1281, 1989.
- Friedmann, *Science*, 244:1275-1281, 1989.
- Fujita *et al.*, *Cell*, 49:357, 1987.

- Fung-Leung *et al.*, *Cell*, 65(3):443-449, 1991b.
- Fung-Leung *et al.*, *J Exp Med.*, 174(6):1425-1429, 1991a.
- Gilles *et al.*, *Cell*, 33:717, 1983.
- Glasgow *et al.*, *J. Biol. Chem.*, 264:10072-10082, 1989.
- 5 Gloss *et al.*, *EMBO J.*, 6:3735, 1987.
- Godbout *et al.*, *Mol. Cell. Biol.*, 8:1169, 1988.
- Golic and Lindquist, *Cell*, 59:499-509, 1989.
- Goodbourn and Maniatis, *Proc. Natl. Acad. Sci. USA*, 85:1447, 1988.
- Goodbourn *et al.*, *Cell*, 45:601, 1986.
- 10 Gopal, *Mol. Cell Biol.*, 5:1188-1190, 1985.
- Graham and Van Der Eb, *Virology*, 52:456-467, 1973.
- Greene *et al.*, *Immunology Today*, 10:272, 1989
- Grosschedl and Baltimore, *Cell*, 41:885, 1985.
- Gupta *et al.*, *J. Biol. Chem.*, 274:7489-7494, 1999.
- 15 Haffter and Bickle, *EMBO J.*, 7:3991-3996, 1988.
- Harland and Weintraub, *J. Cell Biol.*, 101(3):1094-1099, 1985.
- Haslinger and Karin, *Proc. Natl. Acad. Sci. USA*, 82:8572, 1985.
- Hauber and Cullen, *J. Virology*, 62:673, 1988.
- Hen *et al.*, *Nature*, 321:249, 1986.
- 20 Hensel *et al.*, *Lymphokine Res.*, 8:347, 1989.
- Hermonat and Muzycska, *Proc. Natl. Acad. Sci. USA*, 81:6466-6470, 1984.
- Herr and Clarke, *Cell*, 45:461, 1986.
- Herz and Gerard, *Proc. Natl. Acad. Sci. USA*, 90:2812-2816, 1993.
- Hirai, *Hum Cell.*, 15(4):190-198, 2002.
- 25 Hirochika *et al.*, *J. Virol.*, 61:2599, 1987.
- Hirsch *et al.*, *Mol. Cell. Biol.*, 10:1959, 1990.
- Hoess and Abremski, *Proc. Natl. Acad. Sci. USA*, 81(4):1026-1029, 1984.
- Hogan *et al.*, In: *Manipulating the Mouse Embryo: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, 1994.
- 30 Hogan *et al.*, In: *Manipulating the Mouse Embryo: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, 1994.
- Holbrook *et al.*, *Virology*, 157:211, 1987.
- Hollestein *et al.*, *Science*, 253:49-53 1991.
- Horlick and Benfield, *Mol. Cell. Biol.*, 9:2396, 1989.

- Horn *et al.*, *Leukemia*, 15:112-120, 2001.
- Horwich *et al.* *J. Virol.*, 64:642-650, 1990.
- Huang *et al.*, *Cell*, 27:245, 1981.
- Hubner *et al.*, *J. Molec. Biol.*, 205:493-500, 1989.
- 5 Hug *et al.*, *Mol. Cell. Biol.*, 8:3065, 1988.
- Hunger-Bertling *et al.*, *Mol. Cell. Biochem.*, 92:107-116, 1990.
- Hussussian *et al.*, *Nat. Genet.*, 8(1):15-21, 1994.
- Hwang *et al.*, *Mol. Cell. Biol.*, 10:585, 1990.
- Hwu *et al.*, *J. Immunol.*, 150(9):4104-4115, 1993.
- 10 Imagawa *et al.*, *Cell*, 51:251, 1987.
- Imbra and Karin, *Nature*, 323:555, 1986.
- Imler *et al.*, *Mol. Cell. Biol.*, 7:2558, 1987.
- Imperiale and Nevins, *Mol. Cell. Biol.*, 4:875, 1984.
- Inouye and Inouye, *Nucleic Acids Res.*, 13:3101-3109, 1985.
- 15 Jakobovits *et al.*, *Mol. Cell. Biol.*, 8:2555, 1988.
- Jameel and Siddiqui, *Mol. Cell. Biol.*, 6:710, 1986.
- Jaynes *et al.*, *Mol. Cell. Biol.*, 8:62, 1988.
- Johnson *et al.*, *Mol. Cell. Biol.*, 9:3393, 1989.
- Kadesch and Berg, *Mol. Cell. Biol.*, 6:2593, 1986.
- 20 Kaeppler *et al.*, *Plant Cell Reports*, 9:415-418, 1990.
- Kamb *et al.*, *Nat. Genet.*, 8(1):23-26, 1994.
- Kamb *et al.*, *Science*, 2674:436-440, 1994.
- Kaneda *et al.*, *Science*, 243:375-378, 1989.
- Karin *et al.*, *Mol. Cell. Biol.*, 7:606, 1987.
- 25 Karin *et al.*, *Mol. Cell. Biol.*, 7:606, 1987.
- Katinka *et al.*, *Cell*, 20:393, 1980.
- Kato *et al.*, *J. Biol. Chem.*, 266:3361-3364, 1991.
- Kaufman, In: *The Atlas of Mouse Development*, Academic Press, San Diego, 2001.
- Kavanaugh *et al.*, *Current Biol.*, 6:438-445, 1996.
- 30 Kawamoto *et al.*, *Mol. Cell. Biol.*, 8:267, 1988.
- Kay *et al.*, *Hum. Gene Ther.*, 3(6):641-647, 1992.
- Kerr *et al.*, *Br. J. Cancer*, 26(4):239-257, 1972.
- Kiledjian *et al.*, *Mol. Cell. Biol.*, 8:145, 1988.
- Klamut *et al.*, *Mol. Cell. Biol.*, 10:193, 1990.



- Koch *et al.*, *Mol. Cell. Biol.*, 9:303, 1989.
- Komori, *et al.*, *Cell*, 89:755-764, 1997.
- Kondo *et al.*, *Annu. Rev. Immunol.*, 21:759-806., 2003.
- Kozak, *Nucleic Acids Res.*, 15:8125-8148, 1987.
- 5 Kraus *et al.* *FEBS Lett.*, 428(3):165-170, 1998.
- Kriegler and Botchan, *In: Eukaryotic Viral Vectors*, Gluzman (Ed.), Cold Spring Harbor: Cold Spring Harbor Laboratory, NY, 1982.
- Kriegler and Botchan, *Mol. Cell. Biol.*, 3:325, 1983.
- Kriegler *et al.*, *Cell*, 38:483, 1984.
- 10 Kriegler *et al.*, *Cell*, 53:45, 1988.
- Kuhl *et al.*, *Cell*, 50:1057, 1987.
- Kunkel *et al.*, *Methods Enzymol.*, 154:367-382, 1987.
- Kunz *et al.*, *Nucl. Acids Res.*, 17:1121, 1989.
- Lapidot *et al.*, *Nature*, 367:645-648, 1994.
- 15 Lareyre *et al.*, *J. Biol. Chem.*, 274(12):8282-8290, 1999.
- Larsen *et al.*, *Proc Natl. Acad. Sci. USA.*, 83:8283, 1986.
- Laspias *et al.*, *Cell*, 59:283, 1989.
- Latimer *et al.*, *Mol. Cell. Biol.*, 10:760, 1990.
- Lee *et al.*, *Biochem. Biophys. Res. Commun.*, 240(2):309-313, 1997.
- 20 Lee *et al.*, *Nature*, 294:228, 1981.
- Lee *et al.*, *Nucleic Acids Res.*, 12:4191-206, 1984.
- Levenson *et al.*, *Hum. Gene Ther.*, 9(8):1233-1236, 1998.
- Levinson *et al.*, *Nature*, 295:79, 1982.
- Lin *et al.*, *Mol. Cell. Biol.*, 10:850, 1990.
- 25 Lioubin *et al.*, *Genes Devel.*, 10:1084-1095, 1996.
- Lioubin *et al.*, *Mol. Cell Biol.*, 14(9):5682-5691, 1994.
- Liu *et al.*, *Blood*, 91:2753-2759, 1998.
- Liu *et al.*, *Mol. Cell. Biol.*, 21:3047-3056, 2001.
- Long, *et al.*, *Development*, 131:1309-1318, 2004.
- 30 Lucas and Rohrschneider, *Blood*, 93:1922-1933, 1999.
- Luria *et al.*, *EMBO J.*, 6:3307, 1987.
- Lusky and Botchan, *Proc. Natl. Acad. Sci. USA*, 83:3609, 1986.
- Lusky *et al.*, *Mol. Cell. Biol.*, 3:1108, 1983.
- Macejak and Sarnow, *Nature*, 353:90-94, 1991.

- Majors and Varmus, *Proc. Natl. Acad. Sci. USA*, 80:5866, 1983.
- Malynn *et al.*, *Cell*, 54:453-460, 1988.
- Marshak *et al.*, In: *Stem Cell Biology*, Cold Harbor Laboratory Press, Cold Spring Harbor, NY, 2001.
- 5 Marsters *et al.*, *Recent Prog. Horm. Res.*, 54:225-234, 1999.
- Matsuzaki *et al.*, *J. Bacteriol.*, 172:610-618, 1990.
- McNeall *et al.*, *Gene*, 76:81, 1989.
- Mercier *et al.*, *J. Bacteriol.*, 172:3745-757, 1990.
- Miksicek *et al.*, *Cell*, 46:203, 1986.
- 10 Miller *et al.*, *Am. J. Clin. Oncol.*, 15(3):216-221, 1992.
- Miller, *J. Invest. Dermatol.*, 118:216-225, 2002.
- Mordacq and Linzer, *Genes and Dev.*, 3:760, 1989.
- Moreau *et al.*, *Nucl. Acids Res.*, 9:6047, 1981.
- Mori *et al.*, *Cancer Res.*, 54(13):3396-3397, 1994.
- 15 Muesing *et al.*, *Cell*, 48:691, 1987.
- Nagy *et al.*, In: *Manipulating the Mouse Embryo, A Laboratory Manual*, Third Ed. Cold Spring Harbor Press, 2003.
- Nakano *et al.*, *Science*, 265:1098-1101, 1994.
- Nakano, *Trends Immunol.*, 24(11):589-594, 2003.
- 20 Ng *et al.*, *Nuc. Acids Res.*, 17:601, 1989.
- Nicolas and Rubenstein, In: *Vectors: A survey of molecular cloning vectors and their uses*, Rodriguez and Denhardt (Eds.), Stoneham: Butterworth, 494-513, 1988.
- Nicolau and Sene, *Biochim. Biophys. Acta*, 721:185-190, 1982.
- Nicolau *et al.*, *Methods Enzymol.*, 149:157-176, 1987.
- 25 Nobri *et al.*, *Nature (London)*, 368:753-756, 1995.
- Nomoto *et al.*, *Gene*, 236(2):259-271, 1999.
- Nott *et al.*, *Genes Dev.*, 18:210-222, 2004.
- Okamoto *et al.*, *Proc. Natl. Acad. Sci. USA*, 91(23):11045-11049, 1994.
- Ondek *et al.*, *EMBO J.*, 6:1017, 1987.
- 30 Orlow *et al.*, *Cancer Res*, 54(11):2848-2851, 1994.
- Ornitz *et al.*, *Mol. Cell. Biol.*, 7:3466, 1987.
- Palmiter *et al.*, *Nature*, 300:611, 1982.
- Parsons *et al.*, *J. Biol. Chem.*, 265:4527-4533, 1990.
- PCT Appln. WO 94/09699

PCT Appln. WO95/06128

- Pech *et al.*, *Mol. Cell. Biol.*, 9:396, 1989.
- Pelletier and Sonenberg, *Nature*, 334(6180):320-325, 1988.
- Perez-Stable and Constantini, *Mol. Cell. Biol.*, 10:1116, 1990.
- 5    Pesce *et al.*, *Bioessays*, 20:722-732, 1998.
- Pfeffer *et al.*, *Cell*, 73(3):457-467, 1993.
- Picard and Schaffner, *Nature*, 307:83, 1984.
- Pinkert *et al.*, *Genes and Dev.*, 1:268, 1987.
- Ponta *et al.*, *Proc. Natl. Acad. Sci. USA*, 82:1020, 1985.
- 10    Porton *et al.*, *Mol. Cell. Biol.*, 10:1076, 1990.
- Potrykus *et al.*, *Mol. Gen. Genet.*, 199:183-188, 1985.
- Poyart-Salmeron *et al.*, *EMBO J.*, 8:2425-2433, 1989.
- Quantin *et al.*, *Proc. Natl. Acad. Sci. USA*, 89(7):2581-2584, 1992.
- Queen and Baltimore, *Cell*, 35:741, 1983.
- 15    Quinn *et al.*, *Mol. Cell. Biol.*, 9:4713, 1989.
- Redondo *et al.*, *Science*, 247:1225, 1990.
- Reisman and Rotter, *Mol. Cell. Biol.*, 9:3571, 1989.
- Remington's Pharmaceutical Sciences, 15<sup>th</sup> ed., pages 1035-1038 and 1570-1580, Mack  
Publishing Company, Easton, PA, 1980.
- 20    Resendez Jr. *et al.*, *Mol. Cell. Biol.*, 8:4579, 1988.
- Ridgeway, In: *Vectors: A survey of molecular cloning vectors and their uses*, Rodriguez and  
Denhardt (Eds.), Stoneham:Butterworth, 467-492, 1988.
- Ripe *et al.*, *Mol. Cell. Biol.*, 9:2224, 1989.
- Rippe *et al.*, *Mol. Cell Biol.*, 10:689-695, 1990.
- 25    Rittling *et al.*, *Nuc. Acids Res.*, 17:1619, 1989.
- Rohrschneider *et al.*, *Genes Devel.*, 14:505-520, 2000.
- Rohrschneider, In: *Handbook of Cell Signaling*, Bradshaw and Dennis (Eds.), 148(2):147-151,  
Elsevier Sciences (USA), 2003.
- Rosen *et al.*, *Cell*, 41:813, 1988.
- 30    Rosenfeld *et al.*, *Cell*, 68:143-155, 1992.
- Sakai *et al.*, *Genes and Dev.*, 2:1144, 1988.
- Sambrook *et al.*, In: *Molecular cloning*, Cold Spring Harbor Laboratory Press, Cold Spring  
Harbor, NY, 1989.
- Sánchez, *et al.*, *Development*, 126:3891-3904, 1999.

- Satake *et al.*, *J. Virology*, 62:970, 1988.
- Sato *et al.*, *J. Bacteriol.*, 172:1092-1098, 1990.
- Sattler *et al.*, *J. Biol. Chem.*, 276:2451-2458, 2001.
- Schaffner *et al.*, *J. Mol. Biol.*, 201:81, 1988.
- 5 Schwartz and Sadowski, *J. Molec. Biol.*, 205:647-658, 1989.
- Searle *et al.*, *Mol. Cell. Biol.*, 5:1480, 1985.
- Serrano *et al.*, *Nature*, 366:704-707, 1993.
- Serrano *et al.*, *Science*, 267(5195):249-252, 1995.
- Sharp and Marciniak, *Cell*, 59:229, 1989.
- 10 Shaul and Ben-Levy, *EMBO J.*, 6:1913, 1987.
- Sherman *et al.*, *Mol. Cell. Biol.*, 9:50, 1989.
- Singh *et al.*, *Cancer Res.*, 63:5821-5828, 2003.
- Sleigh and Lockett, *J. EMBO*, 4:3831, 1985.
- Sly *et al.*, *Exp. Hematol.*, 31:1170-1181, 2003.
- 15 Spalholz *et al.*, *Cell*, 42:183, 1985.
- Spandau and Lee, *J. Virology*, 62:427, 1988.
- Spandidos and Wilkie, *EMBO J.*, 2:1193, 1983.
- Stephens and Hentschel, *Biochem. J.*, 248:1, 1987.
- Sternberg *et al.*, Cold Spring Harbor Symp. Quant. Biol., 45:297-309, 1981.
- 20 Stuart *et al.*, *Nature*, 317:828, 1985.
- Sullivan and Peterlin, *Mol. Cell. Biol.*, 7:3315, 1987.
- Swartzendruber and Lehman, *J. Cell. Physiology*, 85:179, 1975.
- Takebe *et al.*, *Mol. Cell. Biol.*, 8:466, 1988.
- Tamir *et al.*, *Immunity*, 12:347-358, 2000.
- 25 Tavernier *et al.*, *Nature*, 301:634, 1983.
- Taylor and Kingston, *Mol. Cell. Biol.*, 10:165, 1990a.
- Taylor and Kingston, *Mol. Cell. Biol.*, 10:176, 1990b.
- Taylor *et al.*, *J. Biol. Chem.*, 264:15160, 1989.
- Temin, In: *Gene Transfer*, Kucherlapati (Ed.), NY, Plenum Press, 149-188, 1986.
- 30 Thiesen *et al.*, *J. Virology*, 62:614, 1988.
- Tomic *et al.*, *Nucl. Acids Res.*, 12:1656, 1990.
- Treisman, *Cell*, 42:889, 1985.
- Tronche *et al.*, *Mol. Biol. Med.*, 7:173, 1990.
- Trudel and Constantini, *Genes and Dev.* 6:954, 1987.

- Tsujimoto and Croce, *Proc. Natl. Acad. Sci. USA*, 83(14):5214-5218, 1986.
- Tsujimoto *et al.*, *Nature*, 315:340-343, 1985.
- Tsumaki *et al.*, *J. Biol. Chem.*, 273(36):22861-22864, 1998.
- Tu *et al.*, *Blood*, 98:2028-2038, 2001.
- 5 Tyndell *et al.*, *Nuc. Acids. Res.*, 9:6231, 1981.
- Upender *et al.*, *Biotechniques*, 18:29-31, 1995.
- Van Beusechem *et al.*, *Proc. Natl. Acad. Sci. USA*, 89(16):7640-7644, 1992.
- Vannice and Levinson, *J. Virology*, 62:1305, 1988.
- Vasseur *et al.*, *Proc Natl. Acad. Sci. U.S.A.*, 77:1068, 1980.
- 10 Wang and Calame, *Cell*, 47:241, 1986.
- Watson and Ramstad, In: *Corn: Chemistry and Technology*, 1987.
- Weber *et al.*, *Cell*, 36:983, 1984.
- Weinberg, *Science*, 254(5035):1138-1146, 1991.
- Weinberger *et al. Mol. Cell. Biol.*, 8:988, 1984.
- 15 Weisberg *et al.*, In: *Lambda II*, Hendrix *et al.* (Eds.), Cold Spring Harbor Press, Cold Spring Harbor, NY, 211-250, 1983.
- Winoto and Baltimore, *Cell* 59:649, 1989.
- Wolf *et al.*, *Genomics*, 69:104-112, 2002.
- Wolff *et al.*, *Science*, 246:1464-1468, 1990.
- 20 Wong *et al.*, *Gene*, 10:87-94, 1980.
- Wu *et al.*, *Biochem. Biophys. Res. Commun.*, 233(1):221-6, 1997.
- Yutzey *et al. Mol. Cell. Biol.*, 9:1397, 1989.
- Zhao-Emonet *et al.*, *Biochim. Biophys. Acta*, 1442(2-3):109-119, 1998.

CLAIMS

1. An isolated polynucleotide comprising an *s-ship* promoter capable of promoting transcription operably connected to a heterologous nucleic acid sequence.
- 5 2. The isolated polynucleotide of claim 1, wherein the promoter comprises at least 20 contiguous nucleotides from SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO:5.
3. The isolated polynucleotide of claim 2, wherein the promoter comprises at least 50 nucleotides from SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID  
10 NO:5.
4. The isolated polynucleotide of claim 3, wherein the promoter comprises at least 100 nucleotides from SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO:5.
5. The isolated polynucleotide of claim 4, wherein the promoter comprises at least 500  
15 nucleotides from SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO:5.
6. The isolated polynucleotide of claim 5, wherein the promoter comprises at least 1000 nucleotides from SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:5.
7. The isolated polynucleotide of claim 6, wherein the promoter comprises at least 5000  
20 nucleotides from SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:5.
8. The isolated polynucleotide of claim 7, wherein the promoter comprises about 6.3 kilobases from SEQ ID NO:1 or SEQ ID NO:5.
9. The isolated polynucleotide of claim 8, wherein the promoter comprises about 7.6 kilobases from SEQ ID NO:1 or SEQ ID NO:5.
- 25 10. The isolated polynucleotide of claim 1, comprising a sequence that hybridizes under stringent conditions to the complement of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4 or SEQ ID NO:5.
11. The isolated polynucleotide of claim 10, comprising SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4 or SEQ ID NO:5.
- 30 12. The isolated polynucleotide of claim 1, wherein the promoter is capable of promoting tissue-specific transcription.
13. The isolated polynucleotide of claim 1, comprising a sequence that can hybridize under stringent conditions to nucleic acid segment comprising the complement of i) at least 20 contiguous nucleic acids of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4,

or SEQ ID NO:5; or ii) SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10 and/or SEQ ID NO:11.

14. The isolated polynucleotide of claim 13, comprising a sequence that can hybridize under stringent conditions to nucleic acid segment comprising the complement of i) at least  
5 50 contiguous nucleic acids of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO:5.

15. A nucleic acid comprising a promoter operably attached to a nucleic acid sequence from an *s-ship* gene or a portion thereof and a marker sequence, wherein the *s-ship* gene is disrupted by the marker sequence.

10 16. The nucleic acid of claim 15, wherein the promoter is an *s-ship* promoter.

17. The nucleic acid of claim 15, wherein the promoter is constitutive.

18. The nucleic acid of claim 15, wherein the promoter is inducible or conditional.

19. An expression cassette comprising an *s-ship* promoter operably connected to a heterologous nucleic acid segment.

15 20. The expression cassette of claim 19, wherein the heterologous nucleic acid segment encode a protein.

21. The expression cassette of claim 20, wherein the nucleic acid segment is a reporter gene.

20 22. The expression cassette of claim 21, wherein the reporter gene encodes a gene product that is colorimetric, enzymatic, luminescent, or fluorescent.

23. The expression cassette of claim 19, wherein the nucleic acid segment encodes a therapeutic or diagnostic gene product.

24. The expression cassette of claim 23, wherein the therapeutic or diagnostic gene product is a polypeptide.

25 25. The expression cassette of claim 23, wherein the therapeutic or diagnostic gene product is an RNA molecule.

26. The expression cassette of claim 25, wherein the RNA molecule is an siRNA or miRNA molecule.

30 27. The expression cassette of claim 23, wherein the nucleic acid segment encodes a therapeutic gene product.

28. The expression cassette of claim 27, wherein the therapeutic gene product is selected from the group consisting of a tumor suppressor, oncogene, a cytokine, a cytokine receptor, a differentiation-inducer, growth factor, and a growth factor receptor.

29. A vector comprising an *s-ship* promoter.

30. The vector of claim 1, wherein the *s-ship* promoter is operably attached to a nucleic acid segment.
31. The vector of claim 30, wherein the nucleic acid segment is all or part of an *s-ship1* coding sequence.
- 5 32. The vector of claim 30, wherein the nucleic acid segment is heterologous.
33. The vector of claim 29, wherein the vector is a plasmid, YAC, BAC, or virus.
34. The vector of claim 29, comprised in a pharmaceutically acceptable formulation.
35. A host cell comprising an *s-ship* promoter operably attached to a heterologous nucleic acid segment.
- 10 36. The host cell of claim 35, wherein the host cell is eukaryotic.
37. The host cell of claim 36, wherein the host cell is an embryonic cell.
38. The host cell of claim 37, wherein the embryonic cell is a blastocyst cell.
39. The host cell of claim 36, wherein the host cell is a hematopoietic cell.
40. The host cell of claim 36, wherein the host cell is a stem or progenitor cell.
- 15 41. The host cell of claim 40, wherein the stem or progenitor cell is from tissue selected from a group consisting of skin, a hair follicle, cornea, embryo, gonads, mammary gland, pancreas, and vascular smooth muscle.
42. A recombinant host cell in which one or both *s-ship* genes is disrupted by marker sequence.
- 20 43. A transgenic animal comprising an *s-ship* promoter region operably attached to a heterologous nucleic acid segment.
44. The transgenic animal of claim 43, which is a mammal.
45. A mammal having cells comprising an *s-ship* transgenic sequence.
46. The mammal of claim 45, wherein the *s-ship* transgenic sequence comprises a *s-ship1* coding sequence flanked by loxP sequences.
- 25 47. The mammal of claim 46, further comprising a heterologous nucleic acid sequence encoding a Cre recombinase.
48. The mammal of claim 47, wherein the nucleic acid sequence encoding the Cre recombinase is under the control of an inducible or conditional promoter.
- 30 49. A method for expressing a recombinant nucleic acid in a stem or progenitor cell comprising:
- a) transfecting the cell with an expression cassette comprising an *s-ship* promoter operably attached to the recombinant nucleic acid, wherein the nucleic acid is transcribed.



50. A method of screening for a candidate substance that regulates activity of the *s-ship1* promoter comprising a step selected from the group consisting of:
- (a) contacting a nucleic acid comprising an *s-ship* promoter with an *s-ship* promoter binding protein and the candidate substance under conditions that allow binding between the protein and the promoter and determining whether the candidate compound modulates the binding between the protein and the promoter; and
  - (b) contacting the candidate substance with a cell comprising the *s-ship* promoter operably attached to a reporter gene coding for an expression product and assaying for expression of the reporter gene expression product.
51. A method for identifying stem cells in a population of cells comprising:
- (a) administering to cells in the population a nucleic acid comprising an *s-ship* promoter operably attached to a reporter gene.
52. The method of claim 51, wherein the cells are in an organ.
53. The method of claim 51, wherein the cell are in an animal.
54. The method of claim 51, further comprising sorting cells based on expression of the reporter gene.
55. A method for screening for a modulator of cell function comprising:
- a) transfecting a stem or hematopoietic cell with an expression cassette comprising an *s-ship* promoter operably attached to a nucleic acid encoding a candidate modulator; and,
  - b) assaying the cell for a cell function, wherein a difference in cell function in the cell as compared to a cell in the absence of the candidate modulator is indicative of a modulator.
56. The method of claim 55, wherein the modulator is a candidate therapeutic agent for the treatment of a blood-related disease or condition.
57. A method of treating a patient with a blood-related disease or condition comprising:
- a) transfecting a cell with an expression cassette comprising an *s-ship* promoter region operably attached to a therapeutic nucleic acid; and,
  - b) administering the cell to the patient.
58. The method of claim 57, wherein the cell is a bone marrow cell.
59. The method of claim 57, wherein the cell is autologous.
60. The method of claim 57, wherein the cell is allogeneic.

61. The method of claim 57, wherein the blood-related disease or condition is a blood-related cancer.
62. The method of claim 61, wherein the blood-related cancer is leukemia, lymphoma, or myeloma.
- 5 63. The method of claim 57, wherein the blood-related condition is anemia.
64. The method of claim 57, wherein the blood-related condition can be treated with stem cell replacement therapy.
65. An isolated polynucleotide comprising a heterologous nucleic acid sequence under the control of a developmental decision promoter.
- 10 66. The polynucleotide of claim 65, wherein the promoter is capable of providing expression in embryonic stem cells.
67. The polynucleotide of claim 65, wherein the promoter is capable of providing expression in adult stem cells.
68. The polynucleotide of claim 67, wherein the adult stem cells are differentiated but not terminally differentiated.
- 15 69. The polynucleotide of claim 65, wherein the promoter is capable of providing expression in adult stem cells that are in growing phase.
70. The polynucleotide of claim 66, wherein the promoter is capable of providing expression in a cell from mouse embryonic development stages E3-E18.5.
- 20 71. The polynucleotide of claim 70, wherein the promoter is further capable of providing expression in a cell that is in a developed animal.
72. The polynucleotide of claim 71, wherein the cell is a stem or progenitor cell in the developed animal.
73. The polynucleotide of claim 72, wherein the promoter does not constitutively provide expression in the stem or progenitor cell in the developed animal.
- 25 74. The polynucleotide of claim 65, wherein the developmental decision promoter comprises an *s-ship* promoter region.
75. The polynucleotide of claim 74, wherein the *s-ship* promoter region comprises a sequence that can hybridize under stringent conditions to nucleic acid segment comprising the complement of i) at least 20 contiguous nucleic acids of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO:5; or ii) SEQ ID NO:6, SEQ ID NO:7 SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10 and/or SEQ ID NO:11.
- 30

76. A method for expressing a nucleic acid in a stem cell comprising providing to a cell a polynucleotide including the nucleic acid under the control of a developmental decision promoter, wherein the nucleic acid is expressed in the cell.

5

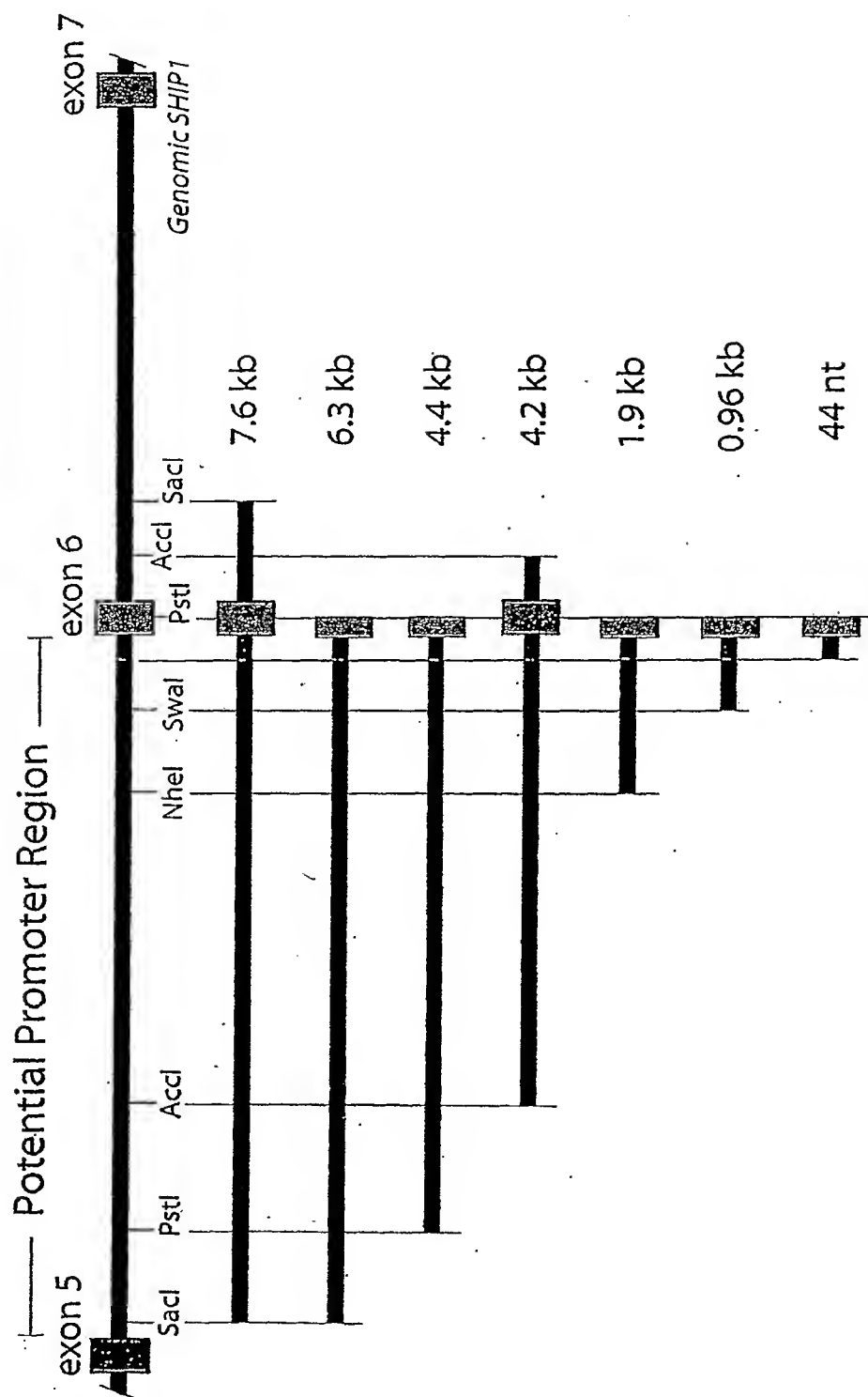


FIG. 1

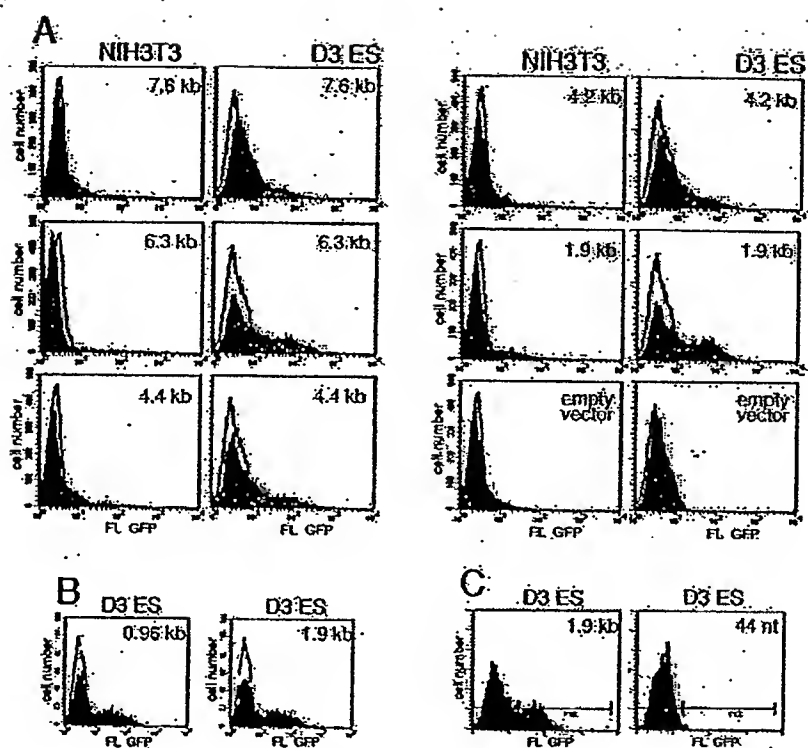


FIG. 2

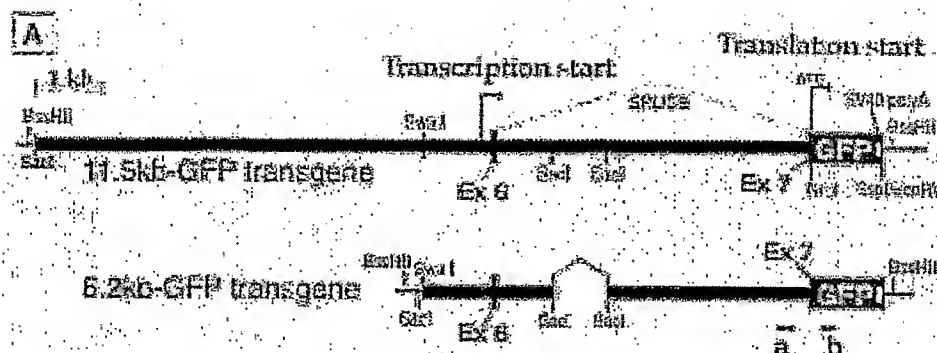


FIG. 3A

## TRANSGENE COPY NUMBERS

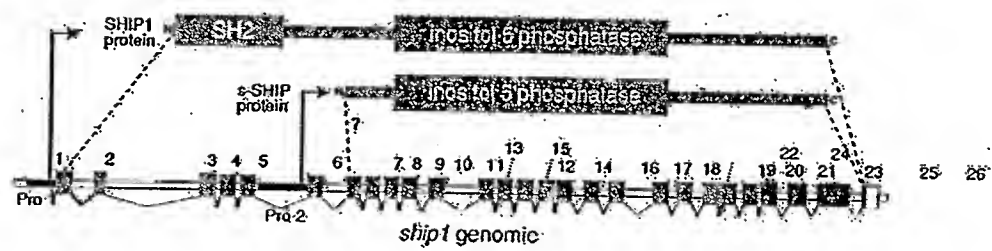
Transgene	Line	# copies
11.5kb-GFP	A	8
"	Q	13
"	Q	11
6.2kb-GFP	Z	17

FIG. 3B

-600 AGAAGTGATCTGGGCTGGACAGACTAGCTGAACTGGCCAGCTCTGGGTTCT  
-550 ATCAAGAAACCTTACCTCCATAAACATAAAGTGTGATGGAGAAAGGCACCT  
-500 AATGTCAACCTCAAACCCCTA Areb6 CCTGCATGTGCACACACATACATCCACAC  
-450 CACACACACACACACACACACACACACCACACACACACACACACAA  
-400 Nkx3.1 Fox factor binding sites ATAAATAAGTAAATAAAATATTTAGCTCTCAGACCAAATCTTGGT  
-350 GAAACCATGCATTTGCATTTGTGTGTGTCTTACAAACACTGAAAGGTTAA  
-300 GAAGCATGCTCCTTAGTAATTTTATAGCAGTTTGCGTTTCCAGATTGAAA  
-250 ACAGATTCTATAGGCTACACAGTGCTAAATGGATTATGCTCAGATACAGA  
-200 TTGAAAAGGATACAGATTGAAAAGGGTCGGGGTCTGGGCCAGGATGACGG  
-150 GCCAACTATCTTTGCCCGGGCTTGTCCTT Stat1/5 CAGGGAAAGGTTACAGGATTCT  
-100 ACCACTGGGGTGTGGCCTATCTGCTGTTAGGACCTGAATTGCCTGGAGTG  
-50 TTTCTAGTTCCCACTAGTTGTTGAACTTTACCTTGAACCTCTGCTCCCAG

Initiator stem-SHIP region

FIG. 4



**FIG. 5.**



## CLUSTAL W (1.82) multiple sequence alignment

```

Hu -----AGCCCAGGATTGGAGTTGTCAATGGATGTGACAATGGGAAAATCGTCT 50
Mo CTCTGGGTTTCATCAAGAAACCCCTACCTCCATAACATAAAGTGTGATGGAGAAAGGCACCT 60

Hu GAGCCTGCATTGTTGGGCTGCTAGGAGGGGATTGTCAT-CAGA-ATCCACAGATCACAGCA 108
Mo AATGTCAACCTCAAACCCCTACCTGCATGTGCACACACATACATCCACACCAACACACACA 120

Hu C-TGGGCAGCCCTA---ATATTTAAAATGCAGATTCTAGACTCAATCAGGCGGGAGCCCA 164
Mo CACACACACACAC CACACACCACACACACACACACACAATAAATAAGTAAATAAATAA 180

Hu GAAATTTGCATTGTTAACACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 224
Mo AATATTTA-GCTCTCCAGACCAAATCTTGGTGAAACCCATGCATTGTCATTGTGTGTGTGT 239

Pax8
Hu TTTATTAACACAGAAGGTTGGGACCATGGA---TAACTAAGTGAAGTCATTTTGTCACT 281
Mo CCTACAAACACTGAAGGTTAAGAGCATGCTCCTTAGTAATTTATAGCAGTTTGCCTTT 299

Hu C-AGATTTGAATT---TTCTACAGGCTATAGAGTGCAGTTTGGCTAAAGCAAAACCTAGG 337
Mo CCAGATTGAAAACAGATTCTATAGGCTACACAGTGCTAAATGGATTATGC-----TCAGA 354

Smad 3/4
Hu TACAGTCAGGACTACACAATTCAGTTCGCTGTGGGTGGGAAGGGATGGCTGGGCCAGT 397
Mo TACAGATTGAAAAGGATA---CAGATTGAAAAGGGTCGGG-----GTCGGGCCAG- 402

P53 motif +/-
Hu GCTGGCAAGCCT-TGATCTTTGCCCGGGCTTGTCTTCTGCGGAGAAATTACCTGCTTCTG 456
Mo GATCACGGGCCAACTATCTTTGCCCGGGCTTGTCTTCTCAGGGAGGGTTACAGGATTCAC 462

STAT 1/5
GATA/Lmo2
Hu CTGGACTGAGG-GTGCCTCATCTCTGGCTAGAGCCCGTGTGCCATGGAAGACTCTTTTC 515
Mo C---ACTGGGCTGTGGCCT-ATCTGCTGTAGGACCTGAATTGCC-TGCAG---TCTTTC 514

Hu CGGTGCCTCACTATTCCTTGATGTTACCTTGT-CCCTGCCCCAG 560
Mo TAGTTCCCACTATGTTGTTGAACCTTACCTTGAACTCTGCTCCAG 560

```

FIG. 6

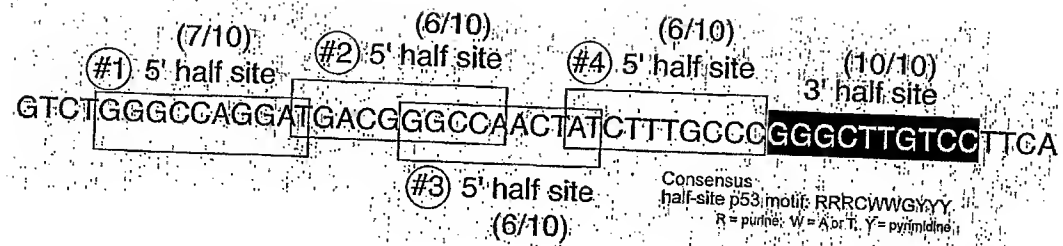


FIG. 7

## SEQUENCE LISTING

<110> ROHRSCHEIDER, LARRY R.  
 <120> METHODS AND COMPOSITIONS INVOLVING S-SHIP PROMOTER  
 REGIONS  
 <130> FHCC:016WO  
 <140> UNKNOWN  
 <141> 2005-03-18  
 <150> 60/554,318  
 <151> 2004-03-18  
 <160> 25  
 <170> PatentIn Ver. 2.1  
 <210> 1  
 <211> 100140  
 <212> DNA  
 <213> Mus musculus  
 <220>  
 <221> modified\_base  
 <222> (27350)..(78168)  
 <223> N = A, C, G OR T/U  
 <400> 1  
 ggcaatttct gagaggcaac aggcggcagg tctcagccta gagagggccc tgaactactt 60  
 tgctggagtg tccgtcctgg gactggctgc tgaccagtc caggagacc atgcctgcc 120  
 tggtcctgg gtggaacct ggcaacatca cccgtccaa ggcagaggag ctactttcca 180  
 gagccggcaa ggacgggagc ttccttgtgc gtgccagcga gtccatcccc cgggcctacg 240  
 cactctgcgt gctgtgagta cccgtctcct cccaactgtc agatccagg accactgagg 300  
 tgtggatcca aagggggaac ccctgtaatg ggagtttgag ttaggtttat gtcataggat 360  
 ggtgggacgt gactggcact tcgttgccct gtggggaggg gagaaggggg ggcagcatct 420  
 gaggccact ttggaccttg gcgttcgagt tcaggagacc tgtgtcatga caggcttgtg 480  
 tgggtgctagg gctctctgag tgactctggg cctcccctat actgcagccc tccatgacct 540  
 gtggtgccag ggtctccgt gacttccgtg ggcaggagca gggacagagt cagggaagaaa 600  
 ctacggcctc tctggtggag ggtgtattgg aatgcatttt ggtcagctca agcgtcagtc 660  
 agggactcag tgaagggcaa ccttggaaca agggctccct cctccacct gctagtatgt 720  
 ggtcttagag ctaattctat ttggggagct gactccgggg tgcatttaat caggatagga 780  
 ttctcgtgt acacatttta ctctaggtg caaggacaca ggaagccaca gactgtctct 840  
 ctgagtaggt ctctgtccct ccttgcactc agctatgtcc ctccctatcc aggtctctgc 900  
 ccccgttggt acccccccac agccagtggt gaagatgttg ctgaatatgc tggttatcct 960  
 aacaacagag gaaacagcca cttgtctgaag gttccttttt aacactccgt cctctagcg 1020  
 tcttctgagg aaggccgcca cccctatagt cctgtggtca gaccctgtcc aggttcagg 1080  
 ctggagcagg gcaggaacac tgtcaggaag ggtgtgccta cttgaacagc acaggtacca 1140  
 tctgatagat tgcacctgga ctgagagaaa gatctttcag agcagctagc tgccccccc 1200  
 ccaatcttca tgcaggaggg aagtgggtga ctgaccacag actgttctga gctctgactc 1260  
 atgttgggac ctgctgtgct aggcatttgt tggctagtgt catcttcaag agagccagga 1320  
 gtgtgggttc tatgaacacc tagacctgcg attaaggacc atgagggtct gacaggttac 1380  
 aaacatcgag gataaggctg gtcatttgct ccagagacc actagcttgt gctgcctacc 1440  
 tctaccacc atcaggctag gcatgccagt ggacttgaat ggaagtaaga ggagctagag 1500  
 tgttcacaga gttgggggca ggggcactag aagcccgtag aggtcacgg ctggtctgagc 1560  
 atcttgtgct ggtggagtta gcagccagct tctgtcacac ccaccagcat atttcaggag 1620  
 aagcactgca ggtggcctgg acctcccaa agcactgtct cttggggaat ctaccagtga 1680  
 gaggccctga aaggaaagag ggcaggaagg tacttttcag tgtgtcaca gctcagcctg 1740

actctactaa	cccagttata	tttttttctg	tcccagtg	atgtgggcca	agggaattc	1800
tggtgggagt	tggtatggg	ctgagagact	gtgtatgca	ttggttacia	atgcacacca	1860
cggatgggt	tctctgtgc	taaactggca	gcacttgga	ggctggagtc	agagcagaag	1920
gagcctgagc	caagaccaag	ggatcttgga	ggaccttcgg	gcaacacaga	ccttgccctt	1980
tttcttcac	tgactccct	gcctcagctg	tcttaagtcc	aagcaaagca	aagatgacac	2040
ggagattttc	aaactaaagg	aatgactacc	acaacctcag	tggtctataa	tgaccccccc	2100
cacacacaca	cacaccctaa	aaatgtagga	aaggcaggac	agtgggtgtg	cacaccttta	2160
atcccagcac	tcgggaggca	gaggcaggca	gatttctgag	ttcgaggcca	gcctgatcta	2220
caaagtgagc	tcaggagacg	ccagggtctac	acagagaaac	cctgtctcga	aaaaaaacaa	2280
aaaaacaaaa	aatgtaggac	aggcatactt	tttaatttga	aaaattgtta	gaagcctgcc	2340
ttttcatgca	aagagactta	acttcctgaa	aaaaacaaac	tttagatcct	tactttctcc	2400
tggtctctgt	ggatgatgga	acctcccca	cctcatcgct	gccccctcgc	catctgcctt	2460
ggccagagtc	caggctcctg	cccaagaaga	taagtcagca	gctttagga	cagcaacaag	2520
gtcgaggtca	gagatggggc	tgtgaaggag	agatggggca	tggtgtacat	gtgggataca	2580
gggcagctga	gttctctttg	gtttcaggag	tgatagattt	ggcacgtgtg	gtgtcttgc	2640
ggacatcagc	cagtctgtgt	gggtcttggg	gcaggcagct	gtgggtcagg	tgccgctggg	2700
actccatcca	tgctccttct	tgtgaccgaa	gggacaccaa	caggggccca	gtgcatgtgt	2760
tggtttgtgt	tgccctctg	ggaagaatgc	agcattgtga	ggagaacctt	cctgctctga	2820
gatttgacta	gacatgacat	gggagagggg	ggtaaatgct	taaagacaag	gttgcaattc	2880
agttccaccc	acgtgacacg	agtgcacatt	cacccacac	attgaccttt	gtttccttca	2940
gagagatcaa	tcctgtcact	tatcaccaag	caagaaggct	tttctttctt	ctcagtcac	3000
attcagagag	ccttggttta	ggggagtctg	gagtacact	ggggccccgg	agcactggcc	3060
tggtggagacc	ttgttagtat	gactggagtt	ctgttctacc	ttccttgaaa	gggaatgtgt	3120
gccttttgag	tggtggcctgt	atcaccttca	cttgagtaga	gcctgtgtca	ccttcaactca	3180
tttgactcca	gttccctcgt	gatgtcagct	cccttcccag	gagcctgtgc	acctgttgt	3240
gggtatttag	ccagggtgat	ggagacctat	taagagtctc	atgagcaggg	acagcgagc	3300
tacaccatgt	gttgagaga	agacaatgct	tctgagggtg	gctcagcaga	tgccgggttt	3360
gtgggtcact	ccagtagctt	cttattgctg	aagctacata	gcaagaattt	gaatgatgac	3420
ccagcacttg	gaaacaactt	gctctttcta	aaagagatga	caaggccaga	ttcaacttgg	3480
tcaagatgac	tggtgtctat	gtgaatggca	tttcccctaa	actactctgg	agtgcctcct	3540
cccttgccag	gagaatatgg	ttgcctctgg	tccagccccg	atgaaggatt	ccctaagcaa	3600
gtgtcttcc	atagctgcac	caggcctgg	ggtggggtaa	gctctgtccc	agggatagaa	3660
tgccaatagc	cttggtagct	ctggcagtg	agaaagaaag	gagaaaaggc	atgggacatt	3720
tacaaccaa	actgccctca	gagaaggcat	tctagtcttt	tgaaagaaac	ggtgtgacca	3780
gacactgggt	gtgataagcc	tgccagggga	gataaaaaca	ggccctgctt	ttaggactac	3840
tggtgagcag	ggtgaagatc	acacacttat	actgtctcac	actgttctt	tggttaggaag	3900
agaactgcag	agaaggagtt	ggtgaggggt	aaaatgccta	gcagagggag	tcaggggcatg	3960
aggctcatct	ccctcccat	cctccattga	aatgtctat	aagggttcca	cagcatgata	4020
atggcttttg	tgcaaacaca	gtgtggcact	gttttccata	tctggcatca	aaggcaaatg	4080
agggaaatac	ctgcataggc	agaaccaga	gctgaaggcc	tatgggctcc	tgagaaacat	4140
gagaaaaggc	ctttgtttga	gaaggatgct	cttgaagact	cattgtgtcc	aagaccagga	4200
ggaagggtcg	aaccaggag	ggcctattta	aggcctattg	tatcatttat	gtaagtggca	4260
gagtgactat	gtttctgccc	ataccagta	ccttgagct	gtcttcccag	gtacagggtg	4320
ggcactggct	taggcgcata	ggttaaattc	acctacaacg	caaggccatt	agcacttctt	4380
aacgatccca	tctctctgcc	tgccacagag	atggaacaag	aactgctatc	gttaccacat	4440
tcattgccag	ccacgtagtc	ccaaggagca	agactctgga	gctgcctgga	atctcttgtg	4500
tcattgtcag	aagacaccta	gaacccagc	tccaaagaga	ggctggccag	accagttcac	4560
tggttttaca	tgccctcagc	tgagggttaag	gtaccactg	ttaagtcacc	catccacatt	4620
ctagtgtgtg	gttcaatgg	ccttagtaat	ggtcacagag	ttacgcaact	gacaccacca	4680
tctaatttca	gagagttctc	attattccca	gaagaactgc	ccccccccc	cacacacaca	4740
catgtgagca	gctgcttgc	tttcttctga	cctctggaaa	gagccagtct	actttctgcc	4800
tttaaggatt	tgccctgactc	ttgcattttg	tgtatgtgaa	attatatggc	gtgcagcttt	4860
tgacatctgg	atctgtccac	ttagcacaat	gccctatgct	cattgaattg	tagcaggcat	4920
ccaatgggtt	gataaccac	tgtgtggaca	caccacattc	cccgggtgag	tggtgttttg	4980
cactgttctt	actttttgac	tggtttgaac	aacagtgtct	gtgaacattc	atcaaccagg	5040
ctctgtgtgg	atgtctgttt	gcaggagtct	tggtggagtga	gaggcagggc	tggtgtcatg	5100
ggtgattttt	atgtttagcg	ttttgaggaa	gtactgaact	cttgtgtctca	gctagagctc	5160
taaccagaca	ttgtgagggt	ccagtttctc	ttctagtta	tctctctctc	tctctctctc	5220
tctctctctc	tctctctctc	tctgtgtgtg	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgtg	5280
tgtgtgtgtg	catgtgtaca	tttgcacact	gtgtgtgtgt	cagecaaaag	acaacttcag	5340

gagttagttc	gctccttcta	ccatgtgggt	cccagggact	gaactcagat	agtcaggctt	5400
aatgacaagt	acccttacc	tctgagctac	ctcactggcc	ctaccttttt	catttcttta	5460
tttaagacag	aggctctcaa	tgccctggaa	cttactgtagg	ttaggtgggc	tgatgttaga	5520
agcaccaagc	ccagtttagt	gtggtcggtt	gcttggtta	gtagtctggg	catcaaactc	5580
aggtgcttat	gctttcaagg	gaagcacttt	atcagctott	acctcccag	cccctcactt	5640
gtttgtgtgt	ttgtgagcta	tgggtctcaca	tagcctaagc	cagcctcaaa	ctccctattg	5700
tagtcaaaat	tggctttgaa	ctcttgatca	tcctgcctcc	actttctaaa	tgtattcacc	5760
acaatatctg	gctgtctttt	tattctaccc	agcttagggg	ctgcaaagca	gcacagcatc	5820
taactgtggg	tttagtttgt	atccctaata	ttaataatgc	taatgggtca	tttgtatgtc	5880
ttacttggag	aactgtctgt	atagtctttg	cacattgaaa	tagctatatc	atttcaatga	5940
taaaatagga	agacaggagt	aatgtggata	ttccatagcc	taacttgaaa	cctctagggtg	6000
ctttgcatat	gtaaaaatag	agctggccat	tagtttttag	agctgaaagc	aaccagtgat	6060
atctcagaag	cagaagacc	atctgtggag	aactttccat	gacaggagag	gagagctgtg	6120
acagtgtcac	ttccgggact	tcctggaggc	ctctgggaga	cacagagctg	tcagtgtggg	6180
cctccacaga	ggaagtgtc	aaagtgactg	aggcaggaag	aggacttcaa	tgacgcaaga	6240
gtgtctgggt	gctgctgagc	cacaggggtt	gacagcctgg	aagcctgggg	agggaggagc	6300
ctgggttagc	acagcaagaa	atgcttagag	gactgggtat	gaagatgaag	tctaggagag	6360
aggctggccc	ccaccactt	ctgacctgag	ctccaactta	gtaaagagat	gccgaaggga	6420
tgaagtcttc	tgatcgctaa	aatgctagct	gttctatggg	aggaaacacc	atgttgggtg	6480
gcacttgccc	tttgagagag	agggtgacgg	agggtgtgag	ttcaatctcc	acagcccaca	6540
tggtgcatgt	ttgtaagccc	agttctagga	agggtggagac	cggaggggtcc	ctaggggttca	6600
tgggcccagc	agactagctg	aattagttag	actcagacca	tgtctcaaaa	caaacaacaa	6660
ccaaaacaac	aacaacaaac	cctagagggt	atctgaagga	ggacatcatg	attgtcctct	6720
gatctctact	cacacattca	tgtatgcaca	catgtgccct	tacacacata	catgggggca	6780
tactgttaac	acataactca	acccatcaat	ttgggtttac	tttgcaattg	aagtaacact	6840
gtgaagggtc	ggdcaaggga	acctctgaaa	aaagagagt	cccacgtggc	cctgcagctg	6900
gaataggcag	tgtagagggt	gacagacctg	gtatgaaaca	gcaaagtctt	ctttcagggt	6960
aacagtagta	ttggctgtgc	ggtggtacac	acaggtgact	gcacactccg	aaagctgagc	7020
aggaagattg	ctgaaagt	cagaccagct	tcccaggcta	catagttagg	ccctatctca	7080
acttaacct	cacagttagt	taagcataaa	gtaaaaagt	taaaaataaac	aaaaattgtc	7140
accacactac	taagagacta	gtgcaagcca	tgtgaatttg	ttgatgtatc	ttctcatacc	7200
catttctatc	acataatata	catacacaca	catacacact	catcaatttt	tctgcctag	7260
tacacaccac	ctttcatttg	tgggtctggg	ttccaacaa	ggcttctctt	ccagtattgt	7320
ttagaatttc	ttccttaaaa	tcactcttgc	tgggaagccta	ccagcaatct	tgtctgaagt	7380
tatgctttat	ggaatgattc	tttgattgac	atttacatta	ccagcaatct	tgtctgaagt	7440
gtggcagcag	caaaaaaggt	cttcagcag	aattgctaca	tatgttactg	gttacaaatc	7500
cccactgcag	ttagagcacc	agcaggaatc	gtcttcaaa	actggacaca	aaatggcaac	7560
ttgctttcca	gaaaaattga	cttaatgtat	gcttctggca	agcagcatga	gaaaaatgcc	7620
acttcttccc	aatatcctat	tctacggaat	gttactttta	agaaaattgg	agtcataaca	7680
tgaagggtca	tctgtgatca	ttttgttttg	cttttgattg	ctgatcctgt	ttgagagggt	7740
tgtgtggcat	tttcttttgt	ggatggctct	ctattctctt	agcttctttt	cttcaattaa	7800
gactgtgtgt	agcttgggaa	gaagggttgt	gataagtccc	aggctaccac	ggactgtata	7860
atgaggtctt	atcttaatat	agggtggagt	ggagctggag	atatggccca	gtgggttaagt	7920
gtttgcttag	tgtgtgtgag	gtcctagggt	caatcccag	taccaaggaa	tgtgtgcaagc	7980
tttgccatct	gcctatgtga	gttctgtcat	gtttgtgaca	aatatatctt	cacgatgggt	8040
tgcttctaag	gttttagattc	tctgacatgg	gagggtcact	gttttctact	gggacttctt	8100
cctgggattt	ctaacacaga	gactccttct	gcacacagaa	atccactcag	aaagggttaag	8160
actgggtctt	tgtagctcat	ccacttgggt	taaggaggga	acctccagca	gtttatcgaa	8220
ttacactggg	ccccagcctt	gtcatctgaa	aaccaagtat	aagaggatct	atgtgttaga	8280
attgttgcat	gtgttgatgt	gtgtatctgt	atcatgtatc	atgatactaa	caaagttagc	8340
caccgtgggt	attattttta	tatgttttat	tgttctactca	atttagcttc	aatagctaaa	8400
tcttcttggg	cttaatatag	gacaagtgat	gaagatttat	gtgtattttt	ctcaagtaac	8460
catccttccc	caacctttta	ttaaatgact	tatgtctttt	ccctccatga	ttcctttgtc	8520
taccatatgc	taagctcttt	ctgggttgggt	tatctgtcca	agaataccca	actacaccac	8580
tacctttaca	aagacataag	cataattttt	caaaaatgtt	tgggggttatt	atctcttata	8640
ttcaagtgtt	ttcccaggta	aaccagatag	taatattttt	caacttggac	aaaacttact	8700
gttatgactt	tcactgggat	ggcaggaaat	cctaattgtaa	gctgtgagct	tgacagtctt	8760
attgtgtctc	ctctctgatt	gtgtaataata	tttcacacag	accctgagtg	tttcttgtct	8820
gggtctctgc	tgtatgactg	tatgtttttg	ttattgttgt	gaatggcaca	tttctcccgc	8880
agttactggg	tcattgataag	cgggttgcctg	catagaattt	tatgtcttca	gctttcattt	8940

tgtcccactc	tttggccact	cctcattcat	ctttcagctg	cttctcctgt	cttttctttt	9000
taaagagatc	ttgtttatctg	tgggtgtgtg	gataccatcc	atgagcacag	atgtgtgcgc	9060
acacatgttt	gtgtttgcct	gcacatgggc	caggagggtga	gatgaggaca	cctgaaatca	9120
tgatccatca	ctctctgcct	tgttcacttg	agacaatgct	tctctctaag	catggggctt	9180
gctgattttg	gctatgctga	ctgaccaggt	ccagcaagcc	tctgtctct	gctaccacc	9240
cacccacact	cagcactgga	tttatagatg	tgtgtgacca	cagctcagtt	ttcttaaatt	9300
ttcagacagg	acctcactat	gtatccttgg	ctggtttaga	aatcacaaag	atgggcctac	9360
ttcccaagat	ccagtattca	aggcacacac	caccacacag	actctctgtc	ctgctttttc	9420
acgtgggctc	cggggatcta	aactcaggtc	tttatgctca	cacagtgage	actcttactc	9480
actgaacgtt	ctaccagcc	caacttttct	gaataggcct	tcatatcatc	tgcaaacag	9540
gatcatagt	tttgcttct	aatgcttaca	accactttct	caaatgccat	agcctattat	9600
attggccagt	atttctaaag	tggtagagt	gcaactgttg	gaaaaaggaa	acgtctgcct	9660
taaggagagc	ttcacatacc	ctgccaccgg	gttgccctttt	ggtcagggcg	ttttattaca	9720
agaatggaat	tgggacatga	gggttctggg	aaggcagata	cctgtcaggg	cctcccatatg	9780
ctgttttgaa	acaatgggtt	tatttaagat	agagatagct	ccttttatatc	ctttgttggg	9840
ttggggccct	cttttgagt	gtgataaggt	ctggtgatct	tttttctttt	tctggcaact	9900
ggggactgag	ggaaacagct	tcttttagaa	tcaaaagcca	ttttgcccac	catatcttat	9960
gaaagagaga	cagacaggca	gacaggcaga	aagaaagaag	aaaaagaaaa	aggaaaacac	10020
caagatgaat	ggggatgaaa	ggaccaccca	aaccaacaag	acacatgtag	tattggggcca	10080
gtggtaaact	cagacagaca	gacagacaga	gggtctccct	gtgagggcct	aactcccgctg	10140
tctttgctta	aagcattagg	tcaagagtga	ccgttgggt	ccctcctgct	gaggagggtt	10200
ctctccctgt	ctctcccca	cctctcacct	gttggccacc	tcaaccctct	tctgtttttc	10260
tctcttttgg	aggggcccaa	ggaagcagg	gatagcctct	gttaatatat	caagtctcag	10320
ccctgtgact	ccagtgttct	ctaggttgg	gtagcaagaa	taaactctct	tgcttgctgg	10380
ggcttaggta	cattttctct	ggttctgtgc	tagccaatat	gccttgctc	cctctttatt	10440
ggtgcatagg	gacagggtca	ctatgtaacc	caggttggct	ttgaactctc	tatactctctg	10500
cttcaacctc	ccaaatgcta	gggttgtagg	tgcacaggac	cataccact	tccatgtgc	10560
tttctgaatg	caatctttcc	attttgactg	ttttctaccc	acctggccca	gaagtgtaca	10620
gcatttatcc	aagaaaagag	aaagcacact	gcaagaatg	ttttctagaa	attagaagt	10680
acacataact	agagtctctg	tgggatttac	aagagcagca	tgtggagcga	gacacagcct	10740
ggactaaagt	ttggactaat	gttttcataa	gggtgatggt	tatgatgctg	tcgaatattg	10800
acagtgtctca	ggcagggtca	tctgagatac	ttctgttggg	ttcagttctg	attttggcct	10860
taatcgtgtg	tgggtgtttc	tagaatatct	tttaaaaggc	caggcaagcc	gggcgtgggtg	10920
gtgcacacct	ttaatccag	catttgggag	gcagaggcag	gtgaatttct	gagttcaagg	10980
ccagcctggt	ctacagagta	agttccagga	cagccagggc	tacacagaga	aacctgtct	11040
caaaaaacaa	aaaaataaaa	ataaaaaataa	aaggccagg	gggctttttt	tectacactt	11100
ggttaaattt	tcttttttgg	gtcagccata	tctttatctc	tctgtccttt	ctttctgtct	11160
gggagcccat	gtgtttcacc	ggctccttag	ggctattcca	taagtagctg	ccctcggtta	11220
aggaaagcca	ttctcacgct	ccctccttgg	cttaggagga	agggtcagag	ctcctgagag	11280
ataaggccag	gtcagatcag	ggcagagatg	gctctacaat	tccctgcct	ggctctgtac	11340
gggttgagcc	taggagccat	aggggcgtat	aagtgggcgt	ggtgaggatt	ctctgtgttc	11400
cagtgatatc	tctgagagcc	gagatgctat	cgggccagat	ggtggtgata	tcgacctggc	11460
ctgtgaaatg	ggtggggttg	gggtggaaca	ttaagttgtg	actagagcca	gaagagcctc	11520
tcaagagaag	ctttcccatg	gggaatgaag	caacgaagtc	atgaactctc	gaaaaagcat	11580
tgccccacat	ttactttaag	tatatatgaa	ttcaaaacag	taaaagatca	taaagctggc	11640
tacatagaaa	aatataaatg	tatacagcaa	cggaagggaa	ccaacaaaca	ttgttgaaat	11700
gaatatcaaa	atccacagta	gcatagatca	gagtcctgtc	tgataataag	cccaggatga	11760
tgtttaaaaa	aaaaacaaaa	aacaaaaagc	aaagagatgg	aaagggattc	aaagaaaaaa	11820
aaaacagaaa	cagccaagaa	caaggacaaa	ccaaaccttt	ctcgtgggga	gaggcaaat	11880
aaggcaagct	gccatagttc	tccccctctc	agatctgaaa	agagatctga	taagccagtc	11940
agcaggctgt	ggggaagtag	acactctcat	atctgttagg	aggaaggaaa	aaaaaccgga	12000
cagactcact	ggagcgcagc	tgtcagttct	tatcaaaatg	taaaatgcc	gcaagacatg	12060
atggctcata	cctgacatcc	tagcacttgg	gagactgagg	caggaggatt	atcttgtgtc	12120
agaagccgac	ctaggctaca	taaaagtgtg	cggtgccacg	caagcatgtt	tgtaaagtgc	12180
ctatgccttt	ttgtctggct	gactctgtct	agttcattgt	ttatccacat	aactagcaac	12240
ccctgacacg	tgtggtgtcc	agtgtgagcc	aggcgtgtgt	ccaaaggctt	tcgatacatg	12300
gatgctttta	gtcttcccag	caacgaatag	atatggcatt	gctgtcatct	ccattttaca	12360
atgaagaaac	aaaaacaaaag	acttcagcca	tgtatgcatg	ggttaaccat	gtatgacaca	12420
cggatgtagc	catgtatata	cacacagatg	ttgcatgtga	gccacatata	catggatgtt	12480
tgttcaccac	aacattctta	cacaggcaaa	acattcaaaa	taactaccac	caccaggctc	12540

tgggtcaaggg	catgacaata	catgtaatgt	ggtagaatgc	tgagaacatg	agcaaaccat	12600
ctgtcatgtg	tgtcttgggt	acttccctat	tgctatgaag	agacaccatg	accaaggcaa	12660
tgtataaaaag	gatatacttg	aattcatgat	catccttagg	cttagaatcc	atgattatca	12720
gggcagggaa	cacggaagca	ggcaggtagc	tgagaactta	catcctgate	cttttacaac	12780
ctcaatgtct	acccccagtc	acacacctcc	ccaacaaggc	caacctccta	atctttccca	12840
gtttttacca	ctgggaacca	agtgtttgaa	cataggagcc	tatgagaaac	attcgcattc	12900
aaaccaccac	aattgtggaag	actttgatga	tattttatga	agaaagcaaa	ctttagaata	12960
acataaagag	taagatttta	tatatttaaa	aaattttaaa	atatatacac	acacatatat	13020
acatatatac	acatacactc	acaatctttt	atgcatgtgt	tttaaaatat	tttataaata	13080
aggataagtc	taagaaaagtc	atggatgata	ggagatataa	tttttttcat	tctatatctt	13140
ctgtgtgtgg	tgcctttttt	caatatataa	tctttttttg	tttgttttgt	gagacagtct	13200
cacatagccc	aggctggcct	ccaacctctt	atgtaactga	ggctaacttt	gaacttttga	13260
ttctcttgcc	tctacctcct	gagtactggg	attattgggtg	tatgtcgcca	ttcttggtt	13320
atgatgggaa	ctgagcttag	ggcttcgtga	atgttaggca	aacactacca	accaagcaaa	13380
tgcacccagc	cctcgatata	aaactttaaac	tgagccaggc	gtggcgcacg	cctttaatcc	13440
cagcactcag	gaggcagagg	cagggtggatt	tctgagttcg	aggccagcct	ggctctacaaa	13500
gtgtgttcca	ggcagccag	ggttatagag	aaacctgtc	tcaaaaaaaa	caaaaaacaa	13560
aacaaaacaa	agaacaaactt	aaactgagtt	tactgttaac	tgtgaaattg	taagttaata	13620
gtaaatgtga	aattgtagt	ttgtgtggca	tgtggaaggg	attagtgaag	tgacttttta	13680
tgaggttgat	ttttatatca	tgggtgtatt	tcctgcatat	atctatgtgc	actctgtgca	13740
tgcagtgtct	gtggagacca	gaagaggatg	ctggattcct	tgggaactgga	attaaagtgt	13800
ggtgcaggcc	tccatgtggg	tgttaggaac	ggaaccagg	tcctcctaga	agatcatcaa	13860
gtgctcttaa	ctactgacat	ttctctccag	cctccctaac	tctcatttct	agattagaag	13920
cattgttgat	agagttttac	agccccact	gcctttgggc	aataggagta	cctagctttc	13980
actgttcatg	agaaccccaa	atagaggaag	acgggtaaga	ggccattcta	gcaggattaa	14040
ggaagacctg	gtgatgggtga	tgttcccaa	aactgtcac	agctcagctg	tggggccact	14100
gaactcctga	ggccaccaac	agacccccat	ggccaccatg	ctaccagata	ccagatggct	14160
ggtgactggt	cacctacggc	cccttcccag	agaatcctta	attttaataa	gttcttccct	14220
tcagagattc	taggaaagt	aaaaaaaaaa	aaaaaggtgg	gggagggggg	ctgggattct	14280
taaaccacaaa	atgtagaagg	agctcttggc	tgggtggcac	ttcctctata	gaaacaacca	14340
gagctctcca	acccagtgtc	tgaaaggggt	aggactcatt	gcaggaagaa	ggaagccatt	14400
aaaacaagga	agccaggaag	agccaaagtag	ggggaatag	gtctcagcag	aagggtgggt	14460
ctggggagcc	tcaccttggc	cgatgtgtct	cttccgagc	tgtagctgag	agctggctcat	14520
catgatacgg	atcttgtctga	gctcctggga	gttttggctc	ttcacacctg	tctcttgtct	14580
cattctcttt	gatgatgtgg	tagctgaggg	gttgaggaa	cccctctaga	tgatggagca	14640
tttagatcag	accacatagc	tatctagcag	tgacgtaaat	ggcatatcca	tggccactct	14700
tggtgtgctg	aggggacagga	acgctgtgct	gtatgcagga	ctgtgaggca	tcagggttcag	14760
ctctgcctca	ccctcatccc	tcagaccacc	gtcatccccg	acatttgcag	gctttgtacg	14820
cagactcccc	cccaagctgt	gtcagacagc	cctgttctat	ctggctctgt	gttttctctg	14880
atggtttgta	atgccgtctt	tgatttat	cagggtccgg	aattgtgttt	acacttacag	14940
gattctgccc	aatgaggacg	ataaattcac	tgttcaggta	agtcccaaac	caacttgtgc	15000
tttccttggg	agaaaagaaa	aaagagaaag	gaaaaacagg	tattctgaat	agttgcaatg	15060
tgagaacctta	gtagagctta	gctttgtccc	ccaaagctat	gtgtctgggg	aactgagtga	15120
ataacagtga	ctcctcatag	atgacagtta	taagatgata	gtcatgcatg	ctaaggatta	15180
ttggtggcca	tggctgagaa	aaggacacag	tgggtgaagga	aagccaagcc	agtgcctatg	15240
ccttgattta	ttcttttgg	tgacttccat	ggatgagcag	atgtcagtgg	cccaggaagg	15300
agcagagggga	ggtgatcttg	caagagagat	gggagattag	atgagatggg	gcctttaaat	15360
tcccatagct	cccttctgcc	caaagcattg	gtcatggcat	ctgaatcact	gtaaaaaggc	15420
caaataatac	aagccataaa	taacttgttt	gattgattga	ttgattgatt	gattgattga	15480
tgagttatgc	cacagtatgt	gtgtgtagag	gtcaaaggac	aacttgaagg	agttgggttct	15540
ctctttctgc	catgtgggtt	ctggggattg	aactcaaatc	ctaaggcttg	ggtgggttact	15600
gccttttacc	actgaggcat	ctcaccagcc	acagctggat	tttaaaaggg	aaacaaacaa	15660
acaacacaa	acaacacaa	aatcatcagt	gccaatcaaa	gccgggggaag		15720
accaccttgg	cagtgtttgt	gtgacaaaca	caggagcttg	tctgaccaca	gagtgtggca	15780
gctgagtgat	gtggaagcca	gagaagctag	tttttgcctg	acttcattgg	ggaagcatgg	15840
gggcgtgcag	aaaggaggtg	acagccccca	tgtcctcagt	agaggagagg	gctgtgtgtc	15900
attaagctca	cagcgaccat	agtaagagtc	acaacagcag	atgggacatc	tgtgtgagca	15960
gtgcagtaag	gaaagacaag	catctagggga	ctgggtcacc	agaacaacaa	ggagcaaaagt	16020
cttcaggatt	ttgctgaaaa	gaaaggggtg	ggaaggggaga	tgtgaggcat	gttggtggct	16080
tccggaatgt	gacaggaatg	atgaattgtt	tcgatcaaa	gggtagactc	agagctgcct	16140

tgttgaaagg	cagtccttag	gcaaaaggcc	ccagaaacat	ggaagcgaga	gagaagttat	16200
gaacagaggc	aaagcctcgg	agggcagctc	atgatggaat	tcaaaagaaa	ccacatcagt	16260
acttccaaca	gtcacttcct	ccagacgggtg	ggactogaag	ggggccagtt	ccatccaggg	16320
tgtggcaaaag	cttcctgcta	ggggaagggt	ctgctcatac	tgagcttcat	gggagaggcc	16380
agcgtgaacc	actgtgggct	ggagtaacgg	ctgtactgca	tagcatgggtg	gtcaggctga	16440
gcacagttca	tactttttct	cgtgtgcgca	tgaacaccag	gatggaaggc	ggggcggtatg	16500
ggaggcaggg	cggatggaag	gtgtgggtcta	ggggtcaggg	agtccctttc	ccagcccagt	16560
cattggtggg	aagcaacaga	acaaatctga	gtaccgcgaa	gcaagcgccc	ttttttaagc	16620
ttggcaggtg	ttcgttgctt	gaatactttc	tgagatgtgg	acgccaccag	cttctggaga	16680
catggctcag	gggagggtggc	ttaggcatgc	agccgtgggtc	tctacttttt	tttttttttt	16740
ttttttttct	aagcctagga	ctgtgtgaac	tttgccctgg	ggcttgggtg	ccttggcaag	16800
tggcttcctg	ctaggagtca	actctcttaa	gttaaaagtt	agtcggcatc	tttgatcttt	16860
ctcccctttt	gtggtttctt	ctgttccttc	aaagtgatga	agactttaag	tatgactgga	16920
aaagagggtga	ctctgcggct	tcttggttag	ggtttgtgcc	cacagcttgt	gaacccataa	16980
ctctgcctgt	ctggagtctg	taccacagtt	ctcttcactc	ttgcttccgt	ggcttctgt	17040
gacccacagaa	atgaccacag	accttcagtg	ggactaggag	cttcagtgtg	ctttgggagt	17100
ccctctggct	gatttttttt	tttttttttt	tgcaaaagtt	tagtgccttt	ggtggcctga	17160
aggacactag	agaagcccag	gccttgggac	cagaggctct	ggggccctgg	ctgggttcta	17220
cctcagttag	tcagctcttc	tctgtgggtt	gctcctcccc	tgctgacagc	actgctcttc	17280
tcagatccca	gagggaacct	gattggctca	gccaccaaca	gctctctctg	cctgaaccac	17340
tcaccagcta	ctgagctctt	tgctgcttgg	ctgttaggaa	gcaggtacct	gcccaccact	17400
gacagatgag	gacagatcat	agcctcctgt	ggctacctca	gccagggcag	tgagcaggca	17460
gtttttatag	aagagatatg	gccttgctgg	acacaattgg	tttcagtctt	gcacttgacc	17520
tatgaccttc	ttggtcatct	ataccagcca	agggacagga	agccatgggc	ctaaatgggt	17580
tgaagtccc	cacctataag	gatgtcttgt	cctataaaaa	acaaacaagc	aaacaaaaca	17640
acaacaacaa	aatccccaca	cattttacaag	tgctggagat	gtagataagc	tcaattggta	17700
gagagcctat	ctagcatgtg	cagacatggt	tcagttccct	agtactgtgt	aacacctggc	17760
atagtagcac	atccctgtaa	tcttagcact	tggggataga	ggtatagagt	taagaggatg	17820
gaaagtcttc	aggtacatag	tgagttagag	atcgccacac	atgcatgtgc	acacgctcat	17880
gcacacacac	atgcatgcac	atgtacagaa	cacacaccag	acatgctaaa	acaaacaata	17940
taaggaacaa	atcaaagctc	acagccagtg	aaggcaaggc	agaaacattc	ttcaggaagc	18000
gttggtatatt	ggccaactgct	tattttaact	ccagtctctc	ttttcatggt	gaagcaaaac	18060
ctgcattctt	tctgagctt	agccacagtt	tctgtataac	tcagtccctta	tacctaaacc	18120
tgcttagtgt	cagcagcagc	aggagcaagg	ctgagactgc	acagcgacac	accaagggcc	18180
ttcatgggtc	cagctcttct	tatgggaggt	ccttgttgaa	acattcattc	caccttcaat	18240
ctttcccaag	gacaatctgt	atctcgtggt	ccagaaggct	cactgatgga	gggaagcacg	18300
acccccacca	ctcagggact	cagctctgac	ttcattactt	atagacctac	aaaatcagag	18360
gtgggaaggg	tggcagcagg	attctcttgg	ttaacgcggg	agacactagc	gttcacaacc	18420
ctgcttccaa	gctaagctct	ctttctcagg	gctcagaaca	tgtttccagg	cataaatggg	18480
tactgaaaca	cttcggcaggt	atgggaccca	agaccatttc	acggagggga	ggcagggtag	18540
tggtctctct	gagatggggc	cttccctggc	tatttggatg	gtatttgatc	atgaatagca	18600
cttggctgtg	tgctctgtgc	attcaggatt	ttgcagccac	atggctttgt	tcagggactg	18660
tgaagagacc	aaacaggaag	atagaatggt	acatctcccc	tgctgtgggtc	acaattctgt	18720
tcacgaatga	cagggatgggt	ctccatggag	gccattgttc	tctttactgc	ccttgttctt	18780
agtaaggaga	catccatgtc	cagttcttgg	taagaaggga	gacagggggc	tggtgagatg	18840
gctcaacggg	taagagtgtc	gactgctctt	ccaaaggccc	tgagttcaaa	tcccagcaac	18900
cacatgggtg	ctcacaacca	tctgtaacga	aatctgtatg	cctcttctgg	agtgtctgaa	18960
gacagctaca	gtgcacttac	atataataaa	taaataaata	tttaaaaaaa	aaaaaaaaaa	19020
aagaagggag	acagccacat	gttcttcagt	ctgcaagggtg	aaaaaagaat	gggaagaaaa	19080
tgggagggaa	agagatgctg	agaacagtat	aaaggggggg	gactgaaaga	cataaaagaa	19140
aaggaataat	gagatgaggg	aaagataatg	aaggaaggaa	cagtcaagaa	agaggaaggc	19200
tggtgacgtg	atgggtacgg	tttctacaaa	ggaaaagagg	tgggtgcatg	cagggcgaga	19260
agctgcaaaag	gtgatggagg	atgaagatgg	aagtgtgctg	tgggctgtgg	aagggtgtctg	19320
tggaggatgt	ctatccctgg	ctctgctccg	ggcactgaag	caaggatcca	cagacactgt	19380
gagacagggt	taacgcagat	tctcagaccc	agagagtgtt	ctgatttaaa	gtocatgcgg	19440
acatccctctg	ggtactgtgt	gaggccagag	gcaggcaggg	aggatgctta	cttagcccat	19500
cgggccccgg	gttcccatgc	ctagaggcgg	tggtcgtgtg	aatcctgggt	ccttcttgaa	19560
tctacaaagg	aagtgactct	cccctcaggg	aagtgcacac	catcttgggc	cgtgacttc	19620
agagtgaagt	gttcacccag	ttgttgctt	ttggagtagc	tttgctccct	gggaagccag	19680
ggagggtctg	aggcaggaca	agaagtctta	acagacaact	gtgttgtgtc	tgaggaggct	19740



gaagagctct	gagcagcctt	tagacatgat	acctctagcc	agggtccacca	ctgaagcgaa	19800
tagggggcctc	tgggcttctc	ttgctgcctg	acacttccta	tgaagaagaa	aaaagcccag	19860
ctcatgatcc	tgtcgcataa	atatctctgt	gaacttcctg	tctctgctga	gccggcagtg	19920
ttggggataa	ggactgtgaa	aactggggta	ctgtgggttt	tcatgcacgt	tcagaagaag	19980
agaggctttg	agatggtctt	tcccacagac	gccaaaggtta	ctcttcacaga	cagaaggaaa	20040
ccccactagc	cagagcctgt	cgtgatctgt	gtagatcaga	cgcgtgactc	aagttcctgg	20100
tccctgggtga	tgtctgtcac	attgctctgg	gcagttcggc	atctggagaa	ggaacaaaac	20160
cacgttttca	gagctggccg	gtcacaagcg	tcggggaggg	cgaacggttg	agccgttgtt	20220
atgaagttca	ggaagcgggt	ggcagctgtt	attgttctaa	gtggtcacca	agtgaataaa	20280
agatttcctc	tgagcctctg	ggcaaaggat	gtggctcaaa	cacacatttc	ctgaggttgt	20340
acaaggaaga	atgcagaaca	ggctcagaag	ctgcatggc	caatgtgaag	acagggagac	20400
ccctttgtgt	cctggcaacc	atattctggg	gcctctgtgt	ttacttttgc	tttattttat	20460
tttattttat	ctgctgtaga	tataggtcta	ttttctaagc	ctctgactat	ctcaggacat	20520
cttgagctg	ggttcatatt	taaaagggaa	attcttgggg	ccaagtctcc	cttcagcatc	20580
ccccagaggt	gtcgttccat	agcctttgaa	aatgtagtat	gaataagtcc	caagcaccac	20640
cccaaggcca	gctcgtgcca	gttgtagtag	agagacaggt	attacctttt	gctcgtgca	20700
tcctttcctt	ctgtggatcc	tgctgagatt	gatcaatttg	caacacaggt	ctttcaagac	20760
tcaaagggtt	ttcttgtttg	actttttttt	tttttttaag	ccagtttata	tgttttattt	20820
ttctcagatt	ctactgagaa	agcccagtc	tgccctcatt	tgccagcctt	ctggaacact	20880
cttactccca	gaaacaaaca	agcaaacaaa	caagcaccac	tcacctgccc	cttctcctgt	20940
gcgcctggag	ggagagctaa	ttgattctca	ggtgtctggg	ggtagggggtg	tttacgctct	21000
gccatctctg	cctctaaatg	gcttatacct	ttctcccat	ctctgtgcac	ccaagccgtc	21060
ttgttgctaa	ttctccatat	cccagaatgt	tatggctctg	tttgtttgtg	ttttcttctc	21120
ctgcataagc	cagagctgac	actaactgtt	gataggctaa	gccaggggaa	cagtactcaa	21180
ttgcagaggg	gaaggactcc	gtaggggggt	cttatcaaaa	agtctttctt	agaaacagac	21240
acaagacaaa	atggctccct	ccctcccaac	tagaatatga	tcattggaat	tgctgctgct	21300
gtttcacaac	cgtgagaaaa	ggtgactgga	gaatgggacc	atgcagacag	cagagcagaa	21360
agataaaatg	aagccaggct	ttgaatgggt	ctattggatt	agctgcatcg	aattaacaag	21420
tctaggagct	ggcctgcccc	tgggcttcct	gttagtgaga	tgactctttt	gtgttcaagt	21480
acatatagga	ttctgaaaca	cgtagctgaa	agtggctcaa	catacaaggt	ttggtctatg	21540
ctagggggga	atctcaaaaa	aggaaaccg	tacacttcag	attgagcacc	gagagtcact	21600
gcacaggtcc	acaggtctgt	gtctgcaacc	gtctctccac	ctctctcctt	agtctggcc	21660
agtgcattgt	gcccggattg	gttctgtacc	atccagacat	ccctaaogat	ttgtctctgg	21720
actcaatatc	ctattgctct	cactcaaate	agaaaaagta	caaaaatcaa	tcagcatgct	21780
agcgacctct	gtgccctgga	aagatgtctc	catttctgtt	tctccatggg	atgatttagt	21840
ttctcatttc	agctttgccc	tttgagagca	ggtgtcattc	agagttgtat	tgagcagatg	21900
agagtctttc	ctcggaggaa	ataccatttc	ccccagtgga	accttgcctg	atcactcata	21960
acatagccga	gggcgtgggt	aagggtggct	ccaaggaaag	tgctggccac	acaaacatga	22020
ggatcagagt	ctagaaaaaa	gaagggtctg	aatcccagtg	cgggtggggg	gggggggggg	22080
agaaatagtt	ggatcccggg	agctccctgg	ccggctaggc	tagcccaatt	ggtaagctct	22140
tggctcagtg	agaccctgtc	tcaaaaagta	agacggaagg	gctggagaga	ttggtgatcc	22200
agcagctgag	agcactggct	gctcttcctg	atgaccagaa	cccacgtggg	ggctcacagc	22260
catctgtatt	totaagttca	aaggctccag	tacctcttcc	ttgttgccct	gctagggagc	22320
aagctcacaa	agcacataga	cacatgcagg	caaaacgccc	atccgcactt	aaaaattcaa	22380
ataaaagaaa	atgtaagtgg	acgataactg	aagtccacct	gcatttgccct	ctggcctcta	22440
catacgcatg	cacacatgtg	tatgcgcaca	cacacacaca	cacacacaca	cacacacaca	22500
tgacacacaca	agaacagact	cctcaccctc	cacctgccag	gtgattattc	aatttttatgc	22560
catggtcctt	attttatttt	tgaggtttctg	aaacctttct	cagcccacag	tgagaatcca	22620
gtagctatca	atggattcgg	tctgcttgca	tcttgagaag	acagaagagc	tgccgggtccc	22680
actggacagc	aggattttgt	ctgccctgct	gcagttgtct	ccacagtgtc	gttactggaa	22740
tccatgectt	ctggcacacg	cttttctccc	ggtggctcaag	ttcttccgtg	ccctgagaag	22800
acttctctcc	ccagtgtctc	actcttgctg	gaactacctc	ctgcatttca	gtactgagat	22860
tccaactatg	ccaccaagct	aacctcagtg	actgtcactc	ctctactgac	cctgttacca	22920
gcagctcttt	catctttcct	tcctgtgggt	tctctctcct	gggtccctcac	gacctcatgg	22980
ctcaaggcta	tggggctcctg	tccaggtggg	tactcccgaa	gccaatgttt	cagagaaggc	23040
tcttagccag	gtcttcggat	cctagcacc	aggatgtctg	gtgcgtttct	ttaaagatgc	23100
catctgtgct	tctgaacctg	tctcctcctc	aggagtgtctg	acgcactgtg	gcatggcttt	23160
cttacagagg	tttaaggact	gaaagagggtg	gccaaataag	cacagagtat	agcataccac	23220
tgtaaccagt	gcatgatggg	gcaaataaac	cagaagtcca	gaggctgggt	agctgccacg	23280
ccaacacagg	gtgatccagg	caaggctggg	gacaccagag	gcagaaaaga	gcgagacagt	23340

ggtgagactt	aagaaccacc	aaagtctctg	gagtttggag	gaaggcagct	ggaatgccta	23400
aggtggaact	tgaaccaaga	ttgagtatcc	agggggctct	gactggcttc	caatcttggc	23460
cacgaacaaa	gtgggagacc	ctagattctg	tatctaggaa	aaattgctgg	cacctgctca	23520
ctggggcgga	gtccttctac	tttgaagaa	ccagggaaggc	agaggcaggg	gatagggtca	23580
tggctggtac	ctggaccact	ggccatgtct	aagagagacc	tgaggacata	ggggtcagag	23640
aggggtcccat	gacttcacct	aggggtctgt	tgccattgag	tctgtttagt	gagcctccat	23700
ggaggcaggg	acttgggata	cactatttca	tttaatcctg	ggaagaggaa	tggtactcta	23760
atggaaacaa	gaccagttag	atggctctcc	agacccacct	gtggctgacc	cgccaccccc	23820
caatgtcgct	ttcagagcct	cttttccagg	cctccccatg	ggaaaccatg	tttgcccaac	23880
acccctctcc	actgtaccct	gttctgtttt	cctaaccatcc	cttagcacta	tctacaacta	23940
gttgccagta	gcaacataac	ccagtagtat	acagtcatac	acattgatta	tcttgtattt	24000
ttggatgcc	gaaatccaaa	aatggctctc	actggctgtg	ctaaaatcca	gatatgagca	24060
gagctctgca	ccatttctga	agacttcaga	aggaaatctt	cctctgtcct	tcaagccagc	24120
aatgggtggtc	aatatgtctc	acagtgccaa	cagaatggcc	ctgtgattat	cctgattaat	24180
tggtatcctt	ttgcatctca	gagcagaaca	tgagcaagcc	tagctcactc	atgctggatg	24240
acagcttatt	cacaggttct	gagatgggga	catgaacatc	tatagaaccg	gatggggaaa	24300
tcattattct	gcctccca	gcagctatgt	tcattgaccta	tccattcagc	tatcctacct	24360
atctccatcc	tgagaccaga	agatggatcc	aagagaactg	agaccttgcc	tgtcttgtac	24420
tgaagtgcc	gaaagccacc	tacactcatt	aggtactcaa	taagtagtga	gcaaatgcag	24480
gctctgcatt	aggtactcag	tgagataat	gtaaatacta	gatataccta	ggcaccatt	24540
aggtactcaa	taatacttga	tacatccaga	taccagttag	gttctcaata	aatactggat	24600
acatgaaaag	gaaggagtat	tttttgtgtt	tccaaatagc	cctgtccatg	ctatttgcca	24660
aggtgctctt	aggacctctc	tgacacggcg	gatttctgag	ttcgaggcca	gcctggtcta	24720
caaagttagt	tccaggacag	ccaagactat	acagagaaac	cctgtctcga	aaaacccaaa	24780
aaaaaaaaaa	aaaaaaaaaa	aaaataggac	ctctctgcag	cagccagtaa	gactcgaatg	24840
gcctgggatg	atgccagggg	cgtggagcat	ctggagcaag	gctaatacaga	cagttgccct	24900
gggctgcccc	ttgaggcagc	tcctctctgg	gctggctccc	agggccaagc	acatccactt	24960
gtctgcctgc	tgctactcct	gtctcccaga	gtaacttgag	ttttacggaa	agtaaagaga	25020
gaagctagaa	gggagggcca	ctcggggatg	taattaagga	taaaaatatg	ccatgtcctg	25080
ggctctggca	acatggaaat	ctagattgct	catgttccca	ggactgtgtt	gtctgtatcc	25140
aagtgggtccc	aaatgtctgt	tctaaattct	tagcagggac	cattttaaag	ttatgtgtgt	25200
gtctgcctcc	ttccttttct	ttttcttggg	atgccaaagaa	ttcaacccaa	ggccttacat	25260
accctaagtg	agtagtactc	taccactgac	ctacacccca	atgccccaa	agtacatcta	25320
aagacagatt	ttgcttggtc	gaagtgggga	gggggaaatg	catatggtaa	gaagatgaac	25380
ttccccgggt	gtgtgtacga	ggagaaggct	tccttaaagt	tagctggcaa	acatgcaagt	25440
aggagaaggc	tgatatttcc	tgtcaccaac	atcttctgtc	cagaaacaca	gcctaaccac	25500
taagcccagc	tttactatc	tccagtcctg	aaggacacct	cgccctcact	gcctgcccag	25560
ggctcagtga	ttctcttaag	ataaagggaag	agtgggtgga	ataatcttga	cctcaagatg	25620
gaaatgattg	cacaggaaga	gcaaggggtg	gggagattaa	tgtggcctgc	gaaaaaaaaa	25680
acggagaaaag	caagctgagg	gatccatgtg	ggcttttgtg	acacccggga	ccactgcaca	25740
ctatagtctt	gctatagatg	ctatatagtc	ctgctatata	gatgtggcta	ccatctgaga	25800
ggtgctgagt	caacatgtgc	tgtctctgct	tttttttttc	tttaaggcaa	gtagaaagca	25860
ggaatcaaaa	ataccaaaca	caagtaaaat	ccctcaccct	ggaggccagt	ttctcttccct	25920
gggaaagtat	cagtgtgaat	acacttctct	caaatgcagc	tgtagaatct	gtaatgtctt	25980
tggtacctta	aaaatggcgt	tgaagcatgg	tgtgggtctc	cagaggagtt	ctttgaaggt	26040
tcattggagga	gagaactgca	aaaaacttgg	caacaggggc	tggggcagtg	gtgtggtccc	26100
cagaaagcac	ataatggtgc	ttggttggtg	tgaatatgat	taaattccca	gcactgggtg	26160
gaagagggac	aggaggagcc	ctggagctta	cccagcctaa	caaaatatte	tttaaaaaaa	26220
ttcactctga	caagaggtct	ggctgcttga	ctcggggata	aaggcacctt	ccacactatc	26280
cttaccacct	gagttttgat	ctctaggatc	cacataaagg	tggaaagaga	gaaccactcc	26340
acaaagctat	tctctaacct	acatacatcc	tacacacaca	cacacacaca	cacacacaca	26400
cacacacaca	cacacgcttt	ctctagtaag	aaagagacct	acatgaagga	atattcattt	26460
ttattgtaca	tttgggtcaat	tcacctcgcc	taaaagctca	tcaagtgagc	ctccccacg	26520
atatctctta	gctcccacca	agtcactctt	gtctcagtat	ctgtaagttt	cctcaagcaa	26580
tggataaaat	ttgcactctc	agtatttttg	tttgcatctt	ctttattaaa	aataacaata	26640
atcccttttc	acatatgttt	tcattcattg	cattcctttt	cttataattt	gcttatttat	26700
ttatttttct	tttggttcatt	tgttttgagat	agagtctcat	ttgtctcagg	ttggctctga	26760
tcttctgct	tctgctcca	gtgctggaat	aacaggcggt	tttctcccat	gctcagattt	26820
aatattttta	tttaaaaaaa	aaatcttttg	gggcgggaga	aatgtctcag	cagctagaag	26880
tacttgttgc	tctttcagag	gatgtatgtt	tgattcccag	caccctgtg	gcagcttata	26940

atcatccatt	tcaggggacc	cactaggcac	atatgtagt	cacatacatt	tgtgcaaata	27000
ggacatccaa	acacataaaa	taaacaaatc	taaaaaacct	ttctaaataa	aaatatgttc	27060
ttgccaggta	atgggtgtac	atgcctttca	tcacagcact	caggaaggaa	aggcaggtag	27120
aactctgagt	tcaaagccag	cctgggtctac	agagtgtgtt	caggacagg	caaggctaca	27180
caaagaaacc	ctacctctga	cctttgtcct	cccaatatac	atataatttag	taaactaatt	27240
tttatttgat	tagtttactt	tttcttttaa	agggttttcc	gtgtgtgtgc	gggagcgtgt	27300
gcgtgtgggt	gtgtgttat	tggtgttata	tatgtgtgtg	gtatgtgtan	nnnnnnnnnn	27360
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	27420
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	27480
ccacagccac	cacagccacc	accctgtgta	caaggctcgg	gtcacatgca	cacaccacta	27540
tgcctccct	gctctctatg	tagttctggg	gatctgaact	cagaaaattt	ataaactaaa	27600
tcactctctt	agcctgtctt	ttttccaga	ttatattatt	atttcacaag	tctatcctga	27660
cttgtgaaat	ttttgtctt	gtttgtttt	gttttctgag	acagggtttc	tctgtatagc	27720
cctggctgtc	ctggacctca	ctttgttagc	caggctgacc	tcgaactcag	aaatctgcct	27780
gcctctgcct	tccgagtgtc	gggattaaag	gcgtgcacca	ccacgcccgg	ccctgaaatt	27840
tttaaacaga	ggctgggata	tagctcagtt	ggtacttgcc	taatgcttgc	tagaacacag	27900
ggagccggcc	ctgtgatcag	tcctcagcat	cataaaagca	gcattggtagc	acacacctgg	27960
aatcccagta	cttgggaggg	agggctggga	cacacagaag	atcagggtca	ttttcagcta	28020
aataggggaat	ttgcagccag	cctgtgcaac	acgggcttct	gttataaaac	aacaacaaaa	28080
tcatttttagc	ttatcagttt	ttctcctgtg	gattctagat	cttgtgactc	accaacctta	28140
attattgggg	ggaaaatagt	attttttatt	ttctaact	tttatcatgt	ttctatgtca	28200
agtgtcttat	tttgacattt	gctttaaaat	ggggactaca	actcttcttc	ccgtgtactc	28260
agcgcctcc	gatagtaaac	agtccccct	ttcctgccga	ctcacgggtc	ctcagcccat	28320
atgctaactg	cccacttct	ctcagatgct	ggatgtagac	tcctcctcct	gttccactta	28380
ttgattagtc	tctacttgca	ataatgcagc	gatgcacatt	ggtagcagt	acattcttat	28440
ggcctgtttc	ggattttctgc	tccagaagga	gactataaat	caagaccagg	gcagcaagg	28500
ggaagaggaa	gtgggtaaat	aagatggggg	tcagcatagg	tcctgtatga	aggttccagc	28560
cagtccactg	gggacttctt	cttaggtaaa	gctgaatgct	aggtgagtgg	tcagaaccag	28620
acaccaggaa	ctctcagggg	ataacctgtg	catcccacaa	tatgtctcct	aaaggagaaa	28680
gacatctgac	tcctaggacc	attcctaccc	cagctgggtc	ttcctctcac	aaggagagtt	28740
ctagccacca	ttctgcatgt	ccaccctaac	ctatgatgta	ctctcccttt	cccaaaggac	28800
cttaggagcc	cagtgtctgca	gagttcactt	cctctcccat	tcagcactct	cctaagagct	28860
cagcttgacc	caccctacct	ctgggaaggg	aaggcccttt	tatttacact	gcgttactc	28920
ttatcttctt	tgtgtgcact	ttttcctacc	togatgcaga	gccccagctt	ggttttctca	28980
taaacttccc	attgtctggg	ctagcatatt	cagatcaccc	agggctctgt	tgtgtctctg	29040
tcaccagaa	gggtggagat	gcctcctagc	tgactggcaa	ccctgtaagc	ccagctctgg	29100
tgtgaacagc	tgatggcatg	gagagacgta	tgtctgttct	gacttgactg	ggatctgact	29160
gagctggtct	ttttctggg	tggtcctcaa	cagctgatga	atgaatctca	caagatgctt	29220
tgcccaaate	tacatagatg	agcttgatgg	tggcaggcag	gtggctgtctg	tgatgttgta	29280
gccatcaagg	cttagcactg	ttggttatct	gacacatggg	tttctttaag	tatggaccag	29340
tgagaaacag	ttcttaaaat	taacagattt	tttttttcca	cataattatt	tgaagttctt	29400
agctcctttg	cagggagagg	gggggggggg	tccaatgagc	ttgggcttgt	tcctccttg	29460
cctaccatct	cacccagga	ctgttctgct	tctgatacag	aaaaggatct	tcttatgggg	29520
gaagggggag	ggacttgtgc	ggtcccatca	cccctgaag	atctgtcggg	ggttaacaat	29580
accaggggag	gcagtcagtt	tctccagtgg	tgtagctact	ggccaggcat	ccatgtcctt	29640
gtcaataacc	tctcatctat	gctctcttat	gaacaaccca	aactaaactc	atcgcatat	29700
gtgtgtgtgc	atgcatgcat	gcgttgtctc	tcattgtgcac	acgcgtgcgc	gaaccacac	29760
tgcataaatg	agagataaaa	gaggggatta	gaaacagaaa	gagatcagca	ggtagtgtgag	29820
ggggacaaca	gagggtcata	ggaggtaaat	gtgaacaaaa	tgtgtgtgtg	tgtgtgtgtg	29880
tgtgtgaaaa	cctcacaata	aagctgtttt	ttctgtataa	ttaatatata	taaactctaaa	29940
agatttaata	aagagcctaa	tccttaaaac	catgtttcac	tgactggatt	ggtccccatt	30000
gcacacccac	tgtcccctag	accgtctacat	gctttgcatt	ttctcataga	ccctgctttg	30060
gcagattgcc	ccgacctgaa	accttaccac	agacgcacat	aacccaccca	aactgtgaac	30120
tgcagagtaa	gccttctcac	ccattgtgaa	agtcacaaact	tcctcgatct	gccttccct	30180
ttaggcctcc	cctcccccaa	ggcccagttc	tcttgtctag	aggtctcttc	cttccctggg	30240
atggagcttg	ggcgtcacag	cagcaggtgt	aactgcactc	gtttgcctgc	cttggatctc	30300
aggctagact	gtgtcttgc	cagttccttg	gagcagcaag	agggctcaga	ggtaactttc	30360
ggttctctta	aaggctgggt	tttatcaaag	tctacctaac	tgaacttctg	gcctccagag	30420
gtctggggcg	tgccatgctg	aggagctgat	ggacccatgc	tatcatgaca	gcttgccaag	30480
cagagaacac	tttgttcttg	tccttttccg	gctcttctca	gcctcagaca	atcccaccac	30540

ccaccacggc	aagctggaaa	cttcctgttt	tgggtctgggc	ctttgccatc	tttaagttgg	30600
aaaatctttg	tccctgatga	atgctccttg	cctctgcat	tgccccagaa	gttattaacc	30660
tttttgtgac	cactgtaatt	tcttcaggat	tcaatgttgc	cttctaatac	tggggtcatg	30720
cagaaggaaa	accaggtata	gacaggaagc	ccagtcggtg	ggagatgatg	ctgttagaat	30780
ggctgatgtc	actccctatg	acacatgagg	tgacaaacat	ctgagagcta	ttctcctata	30840
aggaaaacaa	tgcccccaat	cttctgccag	tgtctgtggt	gtcaatatte	ccaataagac	30900
gcccactaga	atggctgctt	gggattgaag	tcagtgggaa	catgggtaat	aaccctggac	30960
agtgtgatat	aaactcaaat	tggaagagct	gacaaagtcc	tgactctaga	ctttgagcat	31020
cttgatcaga	ccctcatttt	cctagtggg	gtttctgttg	tcattggagaa	acaccattac	31080
caaaagcaag	ttgaggagga	aaggacttat	ttggcttaca	cttcacatc	acagttcttc	31140
attgaaggaa	gccaggacag	gaacttcaac	aggacaggag	cctggagaca	ggagctgatg	31200
cagaggccac	ggaggtgctg	cttactggct	tgtttctcat	ggcttctca	gtctgctttt	31260
cttatagaac	ccaggaccac	cagcccagg	atgacccac	ccacaatggg	cttggctctt	31320
cccccatcaa	ccactaatcg	agaaaatgcc	ttataggatt	gcttacagcc	tgattttatg	31380
gagtcattgt	gtagacaaca	tttaattctca	gtctgggttt	ctaccctgtc	tttgatcatt	31440
caattcccg	ataaaagaca	cacaacatga	ttattttata	taagctttta	atgcactaga	31500
gctgggtaga	tatctaccct	ctaaaccatc	tgaattctact	tcctaccca	taaccccgag	31560
ttgtcacttg	ccatgtttcca	tctgggccac	ctttaactcc	aagtggccag	catggccatg	31620
tttctataat	tcacctaccc	catggtaact	tctccacatt	ccatcttctc	tcttctctct	31680
cgtgggtttc	ctccaagcct	gggaactccc	aatccctgcc	tgtctcaatt	ctgccagct	31740
atagactgta	ggcatcttta	ttcaccaatc	agggataact	tgggggagga	gacaaggtta	31800
cagagctctt	gggtctatgt	gcagattctc	ttgtccctgg	gggcaaccag	gccttgggga	31860
ccagtattta	gcattattat	acctagcaaa	agaccaaacc	tcacagagg	cattttctca	31920
attgaagttc	cctcctctcc	aatgactcta	atcttgtgcc	aagttaactt	aaaaatagcc	31980
cacatacccc	tgtacctcag	aggggtcagg	caagttcaca	gtcacttcca	cctgacttcc	32040
gggtctttca	ggccctaaac	ccttactcag	agtcatttta	attcagtatt	ttctttctct	32100
tttatcacct	ttggggaggg	attttttttt	ttcccaagac	ctcttttaca	gacttgactg	32160
tgtgatcaaa	cttcccattt	ctctgagctt	cggtttatgc	ctctgaaaag	tgggaaggcaa	32220
ctgcccttag	cagatgactt	accccgggct	tcacagcagc	actaactctg	ctctcgcca	32280
agcagatcgt	ataatcctct	tcacagaca	aggagagtag	tgtagcctga	gagcgttaag	32340
tatcacagcg	ttcacactcc	aacagcaaaa	tgccccaatt	agaagcaact	taagggaagt	32400
aagggttact	tgggggtcatg	gttgagagga	tatgggtcca	catgggtggc	aagcaggata	32460
gtgggaccac	aaggcttgcc	ttactgtatc	cacagtcagg	aagctgaagg	agatggacac	32520
ctgtgttcag	ctgcctttct	cctctttcat	tttttttttt	tttataagat	ctagacctcg	32580
ccccatgaga	aatgtctgcta	cctgcattca	gatggtaaat	ctccctcagt	taatcttctc	32640
tggaaattcc	ttgacaaacc	tgttgagaag	tttgtctagt	aggggattct	aagtcctgtt	32700
aatgtggcaa	tgaagattaa	ctacaaaaca	gtagagcctc	gaattaaact	ctcttgacat	32760
ttgacatttg	taaaattctag	caatgctagg	tctattgtcc	ccacaggcta	taattagaac	32820
taaatggaaa	caacttcaac	agtagcaact	cctctcctct	ctccttctcc	ttctccttct	32880
ccttctcctt	ctccttctcc	ttctccttct	ccttctcctt	ctccttctcc	ttctccttct	32940
tcttcttctt	cttcttcttc	ttcttcttct	tcttcttctt	cttcttcttc	ttcttcttct	33000
tctcctctct	ctcctcctcc	tcctcctcct	cctcccgcn	nnnnnnnnnn	nnnnnnnnnn	33060
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	33120
nnnnnnnnnn	nnnnnnnnnn	tttaaatttg	taactactta	tttgacaggg	tctcattctg	33180
taaccagccc	taaatttttt	acaactttct	acaaatgggt	cttttgccct	gggtataagc	33240
actcaaacca	tgaatgggaa	cattgtatat	ttaaactgta	gcaaccctct	gaggcttttg	33300
ttttcaagtt	gtcagtttga	gagccctgac	atgggcaagc	taacagaata	ggactgtgcc	33360
cagaaaaatg	agtgttagcc	cactgcccc	accttcatct	ccattgggaa	ggaagatctg	33420
gggtctccac	acaggctgct	ccaaaacccc	acccaaacta	ccttctcagc	ccaacaacct	33480
tgaattttcc	atcaccttct	tcacctccat	ccactcacac	agaacgaccg	gacactctgt	33540
gactgtgttt	atgttctcag	tgtatccaaa	ttattacagc	tgagcactgc	agatgaccca	33600
aggcagctgc	aggttttata	cagaaacctc	taaaagtctg	taaacctggg	atggttttat	33660
atgatcttta	taaatttgac	actttttaa	aagatctctt	tgaaaaacct	ttccctccat	33720
gtacatgtgt	atgtgcatgt	gtgcacacac	ttgcacgttt	acatgctaaa	aaaatgatct	33780
gcattacata	aactgattct	aataatgctt	ttcataacct	atgttgaaat	aatttttttg	33840
catgtaaaaca	actatgatat	aggccatttc	cagcctgggt	ttttttttta	aataaataaa	33900
taattttatt	tgaagtaaa	ttgacttagg	aaaaatttta	aaatagtaca	aagaattgca	33960
tttttctctc	tcattctccc	actgagcgtt	ttgacgtata	atacactgat	cattctctct	34020
catattgcac	tgtgtttttt	ctgaattatt	ggagaggaag	ggctctgacag	tacctatta	34080
tcactatctc	attcatagcc	atattcagcc	atcaaaatta	ggacattaat	gctggcctac	34140

cgtccactcc	catagctcca	ctcaaatgcc	tccagcagtc	ttagcagcaa	tataaaacat	34200
cctattttgtg	taactatgtg	aactataatt	tatacacata	aaataacttt	cctcctggac	34260
caatgttttag	tcgccaaacc	tgggggtgct	gttgtgtgtg	gccttccttc	atgatatgtga	34320
cttttttttt	tgtttttgag	aacttggtag	taattttgta	ggctgcttct	tggtttgagt	34380
tcttcagggc	tttctctgtg	gttgggtcca	gttctgcatt	tccatttcca	gcaggacttc	34440
tcagtggtgg	tgctaaactt	actactcctg	ctcaaggtga	catgccggct	gctctgcgca	34500
gttggttagag	tacgctctga	gtgtgaagct	aagataacat	ctgaaggggt	ttctctttgc	34560
aaaattagat	atttgttatt	tggggctctta	atattccctc	cctctctccc	tccgtccctc	34620
cgctcctctc	tccctctgct	taggattgaa	accagagtct	cacactattc	tgttcttggc	34680
cactgatcta	tcttcacacc	ctcaggaatt	cagggctgct	ccaggcatca	tagtaagtcc	34740
aaggccagct	tagaccacac	gagtcctgt	caaaaagaaa	ataaatgact	gcagtcaagg	34800
atagcagcac	agcttgaaat	cccagcactc	agaagcacag	gcaagaggat	caggggtttg	34860
agtccagctt	cggctgcata	ataaaaccct	gtctcaaaat	aaacaaataa	aagactgatg	34920
actaggctga	tagctataaa	caacttgaat	gtcaactgcc	tagcataccc	aaggctctga	34980
attcaacccc	taacaaagca	aataaaataa	aactagaaaa	ttatacacac	acacctgttt	35040
tgtggaaata	atccgagact	atctaaatgt	gtcatctctc	aaactctcac	ccacctggct	35100
ttagcatccc	ctggtgaaat	gcgtctgcat	cattaagttac	catgatagct	gccaaatgct	35160
gactgcctct	tctatcattg	gtcccacagt	tatttaattga	catttcgctg	tagggaaagc	35220
ttcccttctg	atatttttat	gtaaggataa	agttgtgaat	tctcatttta	cccagtggt	35280
tgtaactatc	ttcattatcc	atctcaatgt	tcagagtgtg	ctggctgctc	ctcctgctac	35340
acttggtggt	tcctggctct	ttttgacaaa	ttgtcttccc	gttcaaagca	tggcttttct	35400
ccctagaacc	tactagtcca	gtggcccttt	aatatagttc	ttcatgctgt	ggtgaccccc	35460
aatcataaaa	ctaatttcat	tgctacttca	ttaactataa	ttttgctact	gttatacatc	35520
agaaggtaaa	tatttttgaa	gatagaggtt	tgccaaaggg	gtcacaacca	atagtttgag	35580
aacctctact	ctagaaatat	atatctcagg	ttctgctgtg	ccttcccagc	cccagattgt	35640
aggctcttag	ctgaatttaa	gaacctctgg	gtctttgtaa	tggggaaagc	acccaagatg	35700
tgaacagtgc	tcattgctgt	tagtctctgc	cctctcagca	gacagaacca	agaaggtggg	35760
ggaggaaagt	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgtg	tatacaattg	35820
atactgttac	ttaccattca	aatgccaaac	aacagaattg	ctctgcaact	ccctccctag	35880
cctggcttgc	atgcttatgc	ttgttccctt	gcaggtnnnn	nnnnnnnnnn	nnnnnnnnnn	35940
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	36000
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	36060
gacgaaacaa	tgaaactgtg	ctaataacca	gaggaaaggg	agctgggcat	ggtgggtcat	36120
gcctataatc	ccagtacttg	ggaggctgag	gcaggaggat	ctcaagttcc	actctagcct	36180
gagctatatg	gttccagagc	aaaatccttt	ctcagaggaa	gaagggaagg	ggtgagctgt	36240
cttctttggg	gatatccctt	agagaaacca	ctctcagccc	tagaactaaa	gacaggtttc	36300
tgagtctaaa	gcagtcctcc	aaactattct	cagcccaagc	tgcccttttc	cccaggcttc	36360
tgggtctctc	tagctggcca	cgccctctca	gtctattagt	gttctagctg	taccagatcc	36420
ctgtccctgc	ccctttcccc	ttcttccac	cttaccctcc	tcattcaagt	acagactctt	36480
accctgtact	gctgccaaag	tgatgtcatc	ctgaccagca	gtgggtttgca	cctgtaatcg	36540
ccacactcag	atgatgagtt	tgagtcaagt	ctgggtctata	atgtaagaac	ttgaatcaat	36600
caataaataa	aagttaaaat	atcaatccta	catgaatgca	caccccagct	cttgagtgtc	36660
cccaaaagct	ttcaaaaatta	ttgttaaaaa	taatagtaag	ttctacaaaa	tcccagaggc	36720
cccagaaatc	atgcgctctc	cccttcaccc	acatgaagtc	tttttgccca	aggcagaggga	36780
ggtgagtaga	cattgagaag	aaggcagacg	gaaggtagag	gaaacttcct	gatgactgca	36840
gagataggag	agggcagagg	ggcctgtgga	gagagactca	ggagaagctg	ggtcctagac	36900
acactgataa	ctgcataggg	agatggcaga	ccatgctatc	taaaacacaa	aggcctggct	36960
aggccagctg	aactcctgag	ctgccccctg	tgccaccccc	cttccacagt	cccccttccc	37020
cttccccctg	cccttgcccc	gcacagctct	cagatcttct	gtgtaagttg	tatctgtgga	37080
acacccccaa	gcttggactg	agtctagccc	ctgcctgtcc	ctcaatcacc	gtcttttcaa	37140
ctggcagact	cagggctctc	actaagaagc	agaaataata	aacacgatca	ctggcacccc	37200
tctgcccgcg	gacaaatgtg	ttttcacctc	gctttcaccc	tgtgcctaa	gccttgtctt	37260
ctgtctttct	ccttggacct	gaagaggggt	ctttgatatt	gtgtgaatct	gcattggagct	37320
ctttgggttg	ggttccccac	ttcccacttc	ccaccatcta	taaaatctcc	atggtaactg	37380
aatgggccgg	cctcagaatg	gctaagtagc	taaagatcgc	cttgaatccc	tgatcctcgg	37440
tctctgcctc	ccaagggctg	gcattctggg	cctgcagtc	tgctctggga	agtcttttca	37500
acctaaaagg	aaggctaccc	caaccttcag	gagtgccac	ccggagcacc	acacttctgc	37560
tcagcggctg	gcagacttcc	tccctgtgct	ccagtctatt	tcagttctgc	atthagagat	37620
gtcagacagt	gctgacatta	aaactatccc	ctgggatgtt	tcagcttctg	tagggctctg	37680
accgcccagg	cctgctcttt	cctgctcccc	acccccaccc	tgcccccagg	gtattttcag	37740

tcacccccag	gactgcatct	gaaaaagcag	ccagcagctct	gaccaccagc	tgaccttcag	37800
atgaccttag	tcttaaagg	accagtgcc	tccctcctct	tcctcaaac	tggaccccc	37860
cagcctgaag	gccaaacca	cctccatacc	caaaaacctt	tgtgcatggc	tttgtacctt	37920
ctgggatatt	gctccctggt	agatgcagat	gccaaagat	agaacaagat	ttccctata	37980
agatctagct	gggcagtcca	ctccctcca	cctccctgca	ccttttatnn	nnnnnnnnnn	38040
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	38100
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	38160
ggctgtggtt	atgtaccgta	ccagtttccc	ctgtcaacag	acctggctat	aagacagggt	38220
ttttcaaggc	tcttgtgaag	ggggggggtt	ggtaaagtag	gaaccacagt	tatatcccaa	38280
ggcagagtga	gctgtgcatt	caaagcttga	ctccttgggt	tggtgtctat	aaaggccta	38340
taaaaaacat	tgcttctagt	tactctgtgt	agacatatta	actaatcaga	gacaaggact	38400
atatttcatt	tcgtgaacag	aactaatcct	tactttattg	tcttgtgtgt	gacttttgat	38460
tctgaacata	ataaattcat	gtccagtctg	ttgattaagc	aaatcccatg	gagaggctta	38520
tcagcctcaa	gcctgggata	tcccagttac	aagcattatt	ttaaagtagg	aagagcctga	38580
ggctgagggt	ggagcacagt	gaaagggtc	gggccacacc	atgcaccaca	cccgtcatc	38640
ttcagggttct	ctctctcttg	gatcaaactg	cctgccacat	gctatagtta	atctgttaag	38700
gagagaaggga	attcatgcct	gccaaagtga	agaggagagt	caccggggag	caggggcctc	38760
acggtccatg	ggtattgtgg	ggtatcatte	ctaaggggca	cagtgggggt	acatcagtaa	38820
acagctcaga	aggctgttag	aggcccacag	ctaactcgag	acctttgggg	ttccttctct	38880
atttcaaaat	gacttgggtg	atcttttaag	gaaggcctta	gctgcagaat	gaggaacaaa	38940
agctggcggc	ctagacgcta	agaatccaga	gccacgcccc	tctgctggca	gggaactatg	39000
aactaggcct	tgatctgagc	ctcatgttcc	atgtgacaca	caggtgcttc	tctgccatca	39060
tgattgtcct	ctggttcatt	cactcctttg	agtttagcca	gtgagagccc	cagccagggtg	39120
gtaagacagc	gagaggaaga	tagtgaagaa	gcgggggtgg	gtcctctggc	actggctccc	39180
tcctgctggg	agttctcagt	gggctccagc	actatccttc	tgctggggag	ttccaggggg	39240
gtaggggtaa	actgcttccc	tccgctgtta	gctctagaac	tcttctctct	tacctaccaa	39300
actctcccca	aatgccagtt	ttatatggcc	atctgcttcc	tgctaggact	atgactaatg	39360
cctatgacac	ctttcttgac	cacatttttac	attttaaaat	agtcgaaaaa	gcaccacctt	39420
tgggattggt	gggagaaaag	ggggtacacc	aagccccacc	tctcactggt	tctacctoca	39480
agcccaagca	cacagaagg	atctaacccta	gtttttatttc	ctggtgctgt	aataacaaaa	39540
agcaactaag	gggggtggcg	gaggaaagg	ttgtttttgt	tttcaactcat	aattccacat	39600
tacggccctt	tgtaacgggg	agggtgaaagt	ggtagccagg	cagagggtgg	gggtgctgtt	39660
gggtcccagtg	cttgagagag	acagaggcag	gtaaatctct	gagttcaagg	ttagcatggg	39720
ctacaaaagt	agttcaagga	cagccagggc	tacacagaga	aacctatcg	taaaaaaacc	39780
aaaacaaaac	aaaaaggcaa	tactcctctc	ctattccctg	ccccccaca	cacacacacg	39840
ccataggcca	acctgatcta	gacaattcct	cactgagact	ctcttcccag	ataattctag	39900
actgtgtcaa	gttcacaatt	aaaactacaa	attcttgcc	gtcattctga	ttccagtgtg	39960
gtaacccac	agaatcagg	gttgccagg	tccctcagaat	tacaggaagt	acggtagacc	40020
gagatgtagt	atgtagacc	cgctcctgtg	gggagggggg	cctcttaggc	agggtgaatat	40080
gtgcaacatc	tcatgtggaa	ggggacctct	aggctcatta	ctaaaggaca	ttattgttcc	40140
tcctggccca	cagtcagtac	actgccacct	agaagatctt	cccacgtgtt	gggttaggct	40200
tagtggggtt	tattggctac	acagagccgg	ccattcacta	tccactctgc	ctggacatag	40260
ggccttttaa	gcccacttta	tgaagacagc	atttttatact	gtgcctccaa	gagacccctt	40320
tctctggcct	gttgggaccc	atatctccac	catctcctct	gcctctgtct	gctgcctgtg	40380
aagacagggt	gcttctgct	ggccacatg	gcataacaat	tgaaatttat	ctcatgggaa	40440
tttgcttccc	ttaggagggc	ccttgatatc	ttcttgggta	ctagaagtac	aaagcactca	40500
gagctgtccc	tgagccccag	ctcctgggta	gtgaatctat	attatagtgt	ggccataagg	40560
cctttgcaag	gacatttctc	tgtcaaatgg	accagcacc	cacctggcat	ctctgcctcc	40620
tgtctcatcc	ccatctgtct	catcagttgg	ccttgatatat	gtgcatatc	attgcataca	40680
taaataaata	agtggatctt	gttgctctga	ctgagtacca	caaacagggt	cactttttaa	40740
gattaaagtc	caggcatgta	gcttgaagg	aaagcaaatg	cctagcatga	gcaaggctct	40800
agccagacac	ccgtgaatag	aggcatacag	cttatttggc	tcacagttct	ggatgtagga	40860
tcctaggagc	aatgtgctga	cttctatcta	gggagagcct	tcttgctgta	tcataacatg	40920
gcatggggta	tcatatggcc	agacagcaag	ggtgagctga	gattgctttt	cctctgcaac	40980
aagagccact	gacaccatca	cgtaaacaca	cacacacaca	cacacacaca	cacacacaca	41040
cctattacaa	aggctttacc	tccaaacacc	atgtacatat	gaatacaggg	attgcctttc	41100
taataagtgg	gaagacagtc	caaactatcc	cacacttggg	gagttgtgcc	tatgggaggc	41160
agaggaaagc	ctcccttctt	agcagcgccg	ctccaagctg	gctcatgctc	atgcatccac	41220
ctggaggagc	ctgcttcccc	atccttgaca	tttgacctca	accctacact	ccccatcttc	41280
ccggaagcct	gaccaccta	ggtggtcagc	catcctggcc	tgccaggccc	tgctcctctg	41340

cctccatgac	aatgcaagca	tgagcagggg	ccagggaggc	atggctgact	aacagctctc	41400
agcagtcaca	actgcctcca	gctcctgcct	gctccttttg	gggatccctt	tagcttctcc	41460
ttcctctgct	tgtttaatac	acagcctatt	cacccaccat	tgtttacccc	atccccactc	41520
tgccctgccac	ggccaccctg	aacatgtcta	caattcaata	gtgcagcata	gagcctcgtc	41580
tctggggttg	gttctggctt	tttttttttc	ctgctgtata	agataatgtg	catctcagcc	41640
tgacagcccc	ctaagcccca	atcgcaccag	tgttcacaca	tgacgtcctt	aggcactctc	41700
cactccctga	tgagactcct	tccccagatt	cactcaggag	tcagcccatc	tcaaatttct	41760
ttgatatttt	cattctcttt	ccgaatgccc	agtggcattc	gaatcgattc	gaatctcgca	41820
ttctctctaa	ccagccctga	caagaccaca	cacactttac	ctgaagtggg	ttcacaggtc	41880
ttacatctta	tcaaattatt	cttggtattc	cattcttggt	tgcatthgaa	ccactcactg	41940
gagaatattt	tggaaatctt	ggcaggggct	gccactcagg	tggthtaaca	gtgctcacia	42000
gagattgagc	agcgtggggg	tttgaggaga	caggggaggt	tattgaaaag	aataaacaga	42060
tgggtttgct	aagctctcta	taaatttact	aagacaacct	aagccataca	gcatgagtgt	42120
gtatatatgc	tctctcacgc	tctctctctc	tctttgtcac	acacacacac	acacacatac	42180
acacacacac	gcacacgcac	ggatttttgt	gtgtaaaaca	catgagcaca	caagtacact	42240
tggagagcgg	gaagcacgca	gagggaggcc	tggagcccat	atggacagcc	cgagtcacac	42300
gagactgttc	tcaccaaagg	ccgggtgact	cactacagtg	tagaagaccc	agaacccaag	42360
gttcagagac	aagtctgaga	gcttcctagt	tgtagcagct	atttgtctca	ttggtgtaac	42420
aaaatacctt	acaaaagcaa	cttgaggaag	aaacgagtta	atttggctta	cagttcaggg	42480
gatacagctc	acagcaatgg	ggatggcaca	ggacgggaag	cagaggccac	tggccacacc	42540
atgtccagag	tcaggaaacc	aggacgaacg	ctggagtcag	tttgttttct	cctctttact	42600
cagtcctgga	cctcagacca	tgggatgatg	tcactacat	ttaggggtggg	ttttccacc	42660
tcagtaagcc	ccacctagag	aagcccaagg	acttgtatct	atgggtgagc	ttaattccat	42720
caaattggca	agattcctga	gtctgtccca	ttcacctgag	agcataccaa	ccctcccaa	42780
tcaagcagtc	aaaccatcag	gaataggaat	ctaatataga	tgtgactgag	tcaaccacct	42840
gccttccctc	ctgtgaggac	tttctaactc	aaaccatgc	tcctattggt	ctggtttacc	42900
ccccaaagca	tgggtttctt	tcagcaagaa	ggtagctagc	tggccctaaa	gaaatgggtc	42960
agcctctgat	ctctaaacct	gaagatggac	acaagtcag	ttaaaatgat	tggggcagct	43020
ggatgtggca	gctcatgcct	gtaatcccaa	cacttaggga	gctgagacag	gaggcttgcc	43080
atgagtttga	ggccaacctg	agctactatg	tagaacattg	tcttaaaata	aaataagtta	43140
aataatgggt	agacaagagt	gaactatggg	cagacacttg	gccccatgtg	agggatcatg	43200
acccccacaaa	gccgcaggag	actcactaga	gagagtatgt	atctcattag	ctgcagcaat	43260
ttggtctctt	gctctggtct	ggctgtctgt	agggatagcc	atcttagcaa	gggtatatct	43320
aaatggatga	aggggcatct	tgggtggggtc	tgggtctaa	attcagggaa	gaccttgaat	43380
ttcctactgg	agctgcctct	tccaccaaca	tacaaaagag	tcctgggtccc	agagtgtgag	43440
agcagtagag	cagagtcacc	tccatcatca	aaagcagagg	cacaaactgg	tgatgatccc	43500
tgtgtgtctg	ggacaacttt	ctgacttttg	gctgaaaaaa	aaaaaaaaaa	caaacacaca	43560
agaaacatag	ttaaaggctg	gatgggggac	caagtgggta	agagccttga	ctgctcttgc	43620
agaggatcct	gggttcagtt	cccagcccca	catgggtggt	cacaacgac	tctaactcca	43680
gtctaaagga	tatgatgcc	tcttgtggcc	tccatgggca	ctgcagtcac	atagtagaca	43740
aaacaatttt	atatttttaa	aaagtaacta	agtctagcgt	ggttactagt	gcacaccttc	43800
aatcccagga	ctcaagagac	aggaacagaa	aatctctgt	gagttccagg	ttagccaggg	43860
ttccatagtg	agaccacac	aaaaaaaaaa	acaaaaacat	agaattttaa	caaattggaa	43920
aaaaagtaca	agcatgaatt	ttcctatggt	taaaagcaat	aatcaagtgt	gggtggcaca	43980
gcttgtaatc	tttaagccct	tgaggatcta	atgcagagg	gtgggtgcaa	gcttgaggct	44040
acgttgaatt	atatagccag	accccatctc	aaaaagcctt	aactaaatac	agtaaacaag	44100
ccacgtgtgg	tagcacgtgt	tcttctcgg	aggggtgagc	caggaggatg	acataattca	44160
gcttgggtctg	gatagtttag	aggaaacctg	cttcaaaatc	aagtttaaaa	cggcagactg	44220
ttgatgcctg	gtgacacaca	tctgtcatcc	cagcaccacg	gaggcagagg	caggagatgt	44280
aggacggatg	tttcccttagc	ttgtgtacag	tcctcgcttc	aatccccggg	gctgtaaaac	44340
ataagtcatt	aataaacaaa	tagtcggggc	ccgggtgtgc	actcacggta	gggtctatgtg	44400
tgctttgcat	gtgtgagaat	ctgagctcta	ccttcaatac	caaaaataaa	atataatttt	44460
tagaaaagggt	acttggtggg	gggggagcaa	aaatagctct	gttcaccaca	cacacaccta	44520
aacttgtaaa	aatagaactc	ttggcccttt	gagattatag	gtgattttatt	ttttttaaaa	44580
aactgggtatt	ttctaatttg	gatataataa	aaaaatatat	actaatttgt	atagtaaagt	44640
acttatagaa	atcattttta	acagatggag	agaaaaaat	agtttttaaa	atacttctat	44700
ttgcaccttt	gcaaagggct	cctgggagcc	actggctgcg	ggatgtgagg	cgtgagtgga	44760
gagcttttct	gggatattgt	cactcctagt	ctcttcagtc	tctgtctcag	gtggcctttg	44820
tgcttgggtg	acattctagc	ttaaagtgtg	tctgtccggg	gcttctgtcc	tcaggcatcc	44880
gaaggtgtcc	ccatgagggt	cttcacgaag	ctggaccagc	tcacgcactt	ttacaagaag	44940



gaaaacatgg	ggctgggtgac	ccacctgcag	taccccgctgc	ccctggagga	ggaggatgct	45000
attgatgagg	ctgaggagga	cactggtagg	aagggaagga	agggatggca	agtgtgggga	45060
ggtgacaaaa	ggactctgtg	gtccagcctc	acgggtgctt	ctagactaga	gtacctcttt	45120
ggaggttgaa	ggatgctttc	acaggggtcc	cgtaccagat	atctgcattg	aaattcataa	45180
cagtagcaaa	attacagtta	caaagtagca	acaaaataat	cttatgggtg	cggttactgc	45240
aacctgagga	ccgtattaaa	gggttgacgc	actggattag	acagtcttgc	tggcacagaa	45300
atgattggcc	gggaactcat	atgactctga	agtccaggtg	agactagaat	tcctgctccc	45360
cggcttggtg	cctttgtggc	cttagggact	tctgactctc	tgaaccacaa	tcttgccctt	45420
aaagagatta	gataagagta	gcccataatta	atctctcagg	atgtcacagg	accagatgag	45480
cgcgtgaatg	gaacaccatc	agcagtgcac	tccagaaatg	aagttgccat	cagtttcttc	45540
actatagtga	aagttgagag	gctccatgca	cataagttag	agggtagggga	agtgtctctt	45600
catgtaccat	cactaaaggc	aacccaaatg	gtcatgggtg	gacatgccac	aggctattcc	45660
aaaatttagg	atgcccaaggc	aggaaaagca	tgagtttgag	ctagcctggg	ctccatagt	45720
agacaccacc	atctgtaaa	aggcttctaa	aatatgtcgg	agaaaaagag	tttccaagag	45780
tttgtgaagg	ccagcacaga	gacaaggagc	cagtcttgga	gagacaggta	tgcagaaagc	45840
cctgtagcag	gacacccttc	ctttcatctc	aaaggttctc	tgcctgacag	aagcccacat	45900
gttacctcgt	gagacgctca	gcaggtcagt	gagttcccg	caggttcccc	ggggaagca	45960
ttagacatgt	agtgttgaca	tcccaggcct	ggctcctacc	tgtccctgga	gagaatatgc	46020
caggtgtctt	gttatactgt	gggctagctg	tggattgcct	ggcctctgtt	tcaaccttct	46080
ccttcctgaa	gacctccag	agggcagcca	ggagagctgc	ctcaccatcc	tgagggagcc	46140
acagaggcac	atggtgaagc	atgtgtctta	gatggccatg	gtttgaagcc	tgtcttagatg	46200
ggggactgcc	ctgtctcctc	tgcctgacct	ctagagatca	gcagctgagg	tatagctgag	46260
acccaaattc	acaagaaggc	cacaaggcca	catggcagcc	ttgacctggt	tctgtagcta	46320
tgtccttgac	agggcaccac	tctctcttcc	ctcaggtcac	catgcttttt	ctcagcttgc	46380
tctgtcactg	gcaccccat	agaggcaaac	ccttcaatgt	gtcctataaa	ggctctggag	46440
agaatctgtt	tcctgacctt	gacttcacct	ccccaaagct	cctggtcacc	tgcaagaccc	46500
aaagccaagg	gaccccaggt	gcctggacca	aggcatctgg	tccatcttgg	gtcttctcga	46560
gggcctgcag	ttccagggct	caaaatagggt	tcagatctgt	ctcccacagc	cagagggaaa	46620
ggaatgtaca	aatacactga	agaaaaaac	ccttgtttct	gtcacctgtg	tagtgctatg	46680
gctgaccaga	gtggccatca	gggggcaggg	cagggcttgg	gcaggaagtc	ctgagcttaa	46740
atgtcccagt	caccttgact	ttgtgcacat	gaaggacatt	tgaagggtg	ggcaagggtt	46800
acaatctccc	agatgaaagt	caagttgttt	gatattttag	gtctcagaag	ccctttgctt	46860
tctaaatttg	tttctgaggg	tctagaagag	ggaaaggaag	cctgaggctg	gagagactga	46920
agagaaggaa	ggcggaagtg	agagacagac	agacagacag	acagacattg	ggtgggcttg	46980
attttctgcc	cctagtcctt	gactatactt	tctagacccc	gagagccaag	tcactctgct	47040
gatctgcaa	tcaatctgtc	caaggccagc	tatcttgaag	gtgtccctgt	gcttctcaga	47100
accaactgac	ctaggagggt	acttgcctga	cagcatggta	gccagaacaa	caagaccaca	47160
gccaggggct	tcagatctaa	gggtctgaaa	gcaggtcac	ctcaagaggc	tccctatgta	47220
gtcagcatgg	atttaggcag	gaagaatggg	aatagtgggg	cagactggga	cctggaaactg	47280
gggagcctgg	acctctccgg	cttgtggtag	aaatgcggaa	cagtgtctgt	cggacagcat	47340
cctcactgat	ggcatttgct	cctgctgttc	ttttagtaga	aagtgtcatg	tcaccacctg	47400
agctgcctcc	cagaaacatt	cctatgtctg	ccgggcccag	cgaggccaag	gaccttctc	47460
ttgcaacaga	gaacccccga	gcccctgagg	tcacccggct	gagtctctcc	gagacactgt	47520
ttcagcgtct	acagagcatg	gataccagtg	ggtgagtttc	tacttgaggg	gtcccctggg	47580
aagccggcct	atgcctagga	cccagcatct	agcgtccggg	catgctcagg	cctggactgc	47640
ggggagcagc	agacctgctt	attcctagta	aatgtgaaag	ccaatgtctg	tctttctagt	47700
caggactatc	tccttgcccc	caagtcattg	ttgttcttag	tgcatttctg	tattaagatc	47760
gtcacatttt	tcttgtgaga	ttttatgatc	ttcttgaggt	atatttttaca	ttcataactt	47820
ttaggtcttt	ctgccctgta	agtacaagat	tgtttttttg	tgtagaagggt	cactgccgtt	47880
ctgtggctcag	atggcacctc	cacacccacg	gtcggccacg	ttcctgaagt	gtttctcatg	47940
ggaaatcact	ctacagggct	gctcctgcct	gttgacttac	atcccaccac	aggcagctgc	48000
gggaggacct	ctccatgcca	cagccacacc	ttagtctgag	aaatggtagc	cccaaaccct	48060
aggaaaagca	atgtgactat	ctacgtgctt	gctcgctccc	cacatgcctc	ccatggggct	48120
ccaggatggc	accaccaact	tatagcttca	ctaattggagt	acacagtgtt	aagtgcattt	48180
cttagaactc	catcaagtgc	taagagaaag	gctgactgct	aagagccaga	gcccagggag	48240
tacatgggtc	cctgggatata	gtgacagtgg	tgaggggatc	aggtaacaca	gcctaactgg	48300
aaggagacaa	tttgggtccc	ataccaggcc	aaatccaggc	ctgtggggac	cactatctgg	48360
tcttgcatte	ctggccctaa	tcactcacag	aagtcaggca	tgggtgggata	tgtctgccat	48420
tccaacactt	gggaagtcaa	ctggatgttg	tcttatacct	atagtccag	cactgagaga	48480
aggcttgata	ctgaaagatc	aagagttcaa	agccagcctg	ggctacatag	gcagatccta	48540



tctctccacc	ccacccccaa	ccctagaaaa	atataaaaatt	aatactacat	tgacttccca	48600
atggaggcct	gagtgactcc	aacccctcta	tactgtagtg	ctccagggtca	cctttgagag	48660
tttggaacttg	gcatgagggt	cagacctatt	tgttatctct	aggctagctt	gttcttccag	48720
ccctcggtt	acagctgtgt	gaactaagga	gctagagaa	tggccctcca	cgtgtttctg	48780
aggctgtctc	tttaaaacac	aaaaagtcag	atggaggctg	tatgcacctt	tgatcccagc	48840
actcaggagg	cagaggcagg	cagattttgtg	taagtttgag	gccagcttgg	tctacagggt	48900
gagttctaga	acagccaggg	ctacacagag	aaacccctgtc	tccaaaacaa	acaaataaac	48960
aaacaaacaa	aaagaatgac	catttcccag	ccaattacc	cctccaacag	cacccaaaag	49020
tgttctcatc	ctaagaggta	ctgtactctg	tggcctctac	ctgactttga	tcaattcctt	49080
gctctgcagt	cccacttgag	gggactttca	ttatttcttt	gtttctcagg	cttcccagg	49140
agcacctgaa	agccatccag	gattatctga	gcactcagct	cctcctggat	tccgactttt	49200
tgaagacggg	ctccagcaac	ctccctcacc	tgaagaagct	gatgtcactg	ctctgcaagg	49260
agctccatgg	gtaacggaga	gccctgagag	aggggtgggg	gagcttcatg	ggtaatggga	49320
gcccctaccc	accaggagg	atggccacag	caagagaaa	tgctcattag	agtgacctg	49380
ggtctcctct	ctgtccagat	gtctctgcag	cactcacagt	aattggccca	ggtggagtct	49440
ggaatgttcc	aggcttgttg	gaagctcttg	ctctcataga	atctgagctc	taactgagct	49500
gggaaagtgt	atcattttgt	tattcctttt	agggatattgg	gggggcacgg	atgtatgcat	49560
gctgggtatgt	atgcatgctg	gggatgtagg	gctcatgtg	tgctaaatac	atgtgcctat	49620
tctgtaatgt	tttcttgttt	gtttttaact	caaattaatt	agaggcagtt	tctctgttta	49680
aaaaaaaaag	gtggaagaca	ctgccatctc	atttgtgttg	ggacttgaga	tatatatata	49740
tatatatata	tatatatata	tatatatata	atttatttat	ttatttattt	gtgtatgtgt	49800
gtataagaaa	gcgggggaca	ctcatgtgcc	atgggtgtgta	tggggacaca	cgtgtcctgt	49860
gggggaacat	tgtgtcaggg	gaacaagcat	gtgctgaggg	tggggggata	tatatgtgcc	49920
actgtggtga	tttggttggg	ggagacatgt	atgtgctatg	atacctgtgt	gtaggacaga	49980
ggacagactc	acacaggagc	cttcttacct	ctcagcagtc	ctcagcagtc	caagccctcc	50040
aggctgtagc	cactcttcct	ccttgctaca	gtccacatc	cagccaacac	ataaaggctc	50100
tggcaggaaa	taaattaaat	ttgctttgtg	tgtgggtgct	gacggctcaa	gtcttccggg	50160
gatggtaatg	tttaagcaa	gccaacatca	gttctccagt	gccgaagtta	ttgaatgact	50220
gaccaatggg	taacacttag	gattttttaa	aattattttg	attgtacatt	ttgaattaca	50280
ttaatcttaa	aataaactac	aaacataaac	acactgtggt	tggagctggg	catagtggca	50340
catctcttta	attccagaac	ttgggaggca	gaggctggca	tatctctatg	actttgaggg	50400
cagcctgggtc	tatataatga	gttccaggac	agagagatcc	tgtctcacia	acagagaaaa	50460
ccccataact	atacttttat	tgtactgggt	ggctctacagt	gtaaagaatt	ggcaagaagt	50520
atgaaagatt	acactgggaa	gaaactgaaa	gccatccaga	gtaatgaagc	aaacccctca	50580
caaagtgtggg	aaacatttca	catgggtgag	ggcttgacgc	cagcctgtgc	taattttatgt	50640
tttggtagct	gagcacttag	agcagtttcc	agttctctgg	tgctctacca	atcttagctt	50700
ggtaaatatt	cagggatgag	tcttccatca	ccggtaatct	aatttgccat	tctattttaa	50760
ggctcttaaa	ggcacaggca	gtgcattggg	taaatgtggc	aaataatttc	tttgacaaat	50820
cgaccaattg	tcagattggc	ctgctagtca	tttgtttcaa	tgagaactgt	tttttctcaa	50880
aggatgctct	tgtacaccgg	ctagaagcag	ggctgtcatt	tttataggct	tctgtgggat	50940
tttgtgttg	ttgttctgtt	gttgcaaatc	atttatcact	gagggaaaat	acacacaaag	51000
ggccctttct	ttaaaagtat	acatgtatca	ttttgtgaca	gctcataaga	agctgttttt	51060
ttctgctgg	acacaggccc	tgacctgtgc	tgtgtccttg	ctaagctttg	tcagaccctt	51120
ccacagctcc	ccccaaacac	gagttcccca	gtacctgcct	cacctcatca	ctatggtgac	51180
agcagcctct	gatgcgctg	actctctggc	acattatggc	agtgttaaaa	gcttccatct	51240
ctctctttgc	tgaataatga	acctcagggt	cttcaagaga	ccggaatgtt	cttcacctgc	51300
ctgcacacat	ctcttcactt	tcttttatag	atcaggtagg	gactgggcgt	gtagatggaa	51360
caaactgttt	tccgttcccc	agccatctct	gcagggtgcac	tccacataaa	tcaagtgtta	51420
aaagtgtctt	gattaaacag	gacaggcgcg	ttcttgagtt	catctgttca	catactgtct	51480
ggcaagcgct	gactgagggt	ctcctctgta	ccctgttctg	agaactaaca	aaagacgaat	51540
caacatacag	aaaactgtta	tttagtgact	gattaaacta	acgaaggcat	gggctggaga	51600
aataactcag	cagttaagaa	catttgctga	tcttgacagag	gacctgggtt	gggttccctag	51660
cacacacagg	gacagttcca	gtcccggtgt	gtccttttct	cacttctgtg	gacacaagtt	51720
ttacacatag	tgcacacaca	tacactcaca	tatataaaac	agaacattta	aaagtatggt	51780
taaataacgg	aatcatttat	ataggttttc	atttacatag	gtaaaataggc	aaaaatctgc	51840
attttattgt	ttctaagttt	taattttatt	ttctctgtgt	gtacatacgc	atgcctcctt	51900
atctgtatgt	gcgtgcaactg	tgtgcatgca	tgaacccaca	gagaccagaa	gagtaccaca	51960
gatttctctg	agctggagtg	attgataggc	tgttgggagc	cactccacat	ggggattcag	52020
agttgaactt	cgttctctgc	aagaacagcc	agctcttaac	tgatggcttt	tacctccagc	52080
caccttttcc	tcatttttaa	aatttctctc	cttctttttt	gagacagggt	ctcaataactt	52140

agctcatccc	aacttgaccc	cactcttctc	ttgccttagt	caccacaatg	tttagtttat	52200
aagcatgCGT	cactatgccc	ggctttaaat	aaactcacc	ataatcccag	cactgaagta	52260
gacaaaaggg	aggatcgatg	gggtgactg	gccacaagcc	gtgcttcaag	ttcaatgaag	52320
accctgtctc	aagggaataa	ggcacagagg	atagagccat	acgcctgacc	tcctcctctg	52380
gcctctaccc	aggcacatgt	gcatacacac	accacacaca	cacacacaca	cacacacaca	52440
cacacacaca	gagagagaga	gagagagaga	gagagagaga	gagagagaga	gagaaaacttt	52500
ttcctctttt	tttttaaaaa	tattatttat	ttcatgtata	tgagtacact	gttgctgtct	52560
tcagacaccc	caaaagaggg	catcagatcc	cattacaaat	ggttgtgagc	caccatgttg	52620
ttgctgggaa	ttgaactcag	gacctctgga	aaagcagtca	gtgctcttaa	ccatctcttc	52680
agccccacaa	agaaactttt	aatgagcaaa	taattgcttc	caagtaaata	ctactaatat	52740
atcttctaacc	atactataca	aggaattatt	aaagaacgga	taataggaga	ataaaaaatt	52800
ataagtcaact	ttataatgct	atctaatacca	tctagaacaa	aaacactgta	ataatgcaaa	52860
agagcgcgagt	gcctagatta	aataaataaa	atgcagacca	ataagtaaac	tttatagcag	52920
cacatggaaa	tgacgaaatt	cctaacaaaa	agctcaagat	gggcagttta	tttaaagtga	52980
aatacaggag	aaataaagca	cagaaagata	ctcaagggca	tagaagttaa	catagggggg	53040
ctggcgagat	ggctcaacag	gtaagagcac	cgcactgctc	ttccaaaggt	cctgagttca	53100
aatcccagca	accacatggt	ggctcacaac	catccgtaat	gagatttgac	tcctctctct	53160
gggtgtgtctg	aagacagcta	cagtgtactt	aacataataat	aaataaataa	aactttaaaa	53220
aaaaaattaa	agaagttaac	atagaagccc	actcaggacc	ccactcagtc	ctagagtatg	53280
acattatttat	ggacattaaa	aagagaaaat	tcagcagtag	tgtgcatgca	ctgcatatat	53340
acacaaatcc	ttgagtttca	taccaaattgc	ctttagacca	cttgtggctc	tgcaaacctg	53400
taatcctagc	acttgtgaaa	aggtcagctc	aggaactttg	ggaagggtcat	gaaactcttg	53460
cccctccaga	agggagaggc	taattaacat	ttctcagacc	acagggcggg	aaccgacctg	53520
cgggtgggga	cagactgttg	cccatttcca	gactagggaa	gtccttgtca	cctcattccc	53580
taaagaccaa	tcaattttaa	gggtgcactg	ttccggcaat	catattgtgc	ctagttgtcg	53640
atgctctatt	ctgcccttag	aaaccgtata	aaaactagcg	aaggggtacc	aggggtaacc	53700
ccctctcctt	caggtctggg	acaatcccac	tacactggaa	caataaattc	ctcttgcttt	53760
ttgcattgat	cacagctcca	cttcgtggta	agctaagact	ccctggagtc	ttacattggc	53820
aaatgcaggc	aaaagaatcc	gaaactcaag	gtcatctaaa	actacatagc	aagcatgctg	53880
ctagcctggg	ctccatgaga	ccctggggga	ggggcagagg	gagaccgttc	agaagacagt	53940
caagatgttg	cagcagcaca	ggcagcctgg	ccaccagtgc	tgtcaccaga	catgttaatg	54000
ttggaataaa	gcctcaatca	tgactctccc	agttttataa	ttggaaataa	gaaaggaaaag	54060
actataggaa	caactgtgtt	cagaacacta	tttataatag	caaagatctc	agagtaaccc	54120
aaacttctag	acattgattt	gggaagatct	cttggcagct	tattttgaaa	actttacaat	54180
gttaaatatg	taaaaacaag	gacagttttg	ttttttgttt	tgttttggtt	tgttttaggg	54240
atatatatcc	atatatgtat	atgaatgaaa	acccaaactt	aaaattcccc	actatgcttt	54300
aaaggctttc	tgacaataac	agaaagagaa	atagagaatc	cataaaaact	agttctgaaa	54360
ctatcaatag	gcttgacact	ctttagctgc	caggagagct	gaatctgaac	acagggaaacc	54420
ccaccagca	ccccaaattt	ggattattgt	tttattttat	ctttccccta	cccccaagac	54480
agggtttctc	tgtgtgggtc	tggtgtgctc	ggaactcgga	gatectctgc	ctctctgctc	54540
ctctgctctc	ctctctctct	gcctctctct	gcctctctct	ctctctctct	ctgctctctc	54600
ctctctctct	gcccctctct	gcctctctct	gcctctctct	gcccctctct	gcccctctct	54660
gcccctctct	cttcccctct	ctgctctctc	ctctgcccct	ctctgctctc	ctgctctctc	54720
ctcctgagtg	ctgggattta	aaggcatcag	ccatcacttc	cagcttcctt	tatcatttta	54780
aaaagaattt	cctatgtgac	tactgtattt	aaatcaccac	acggccaata	ctcccccca	54840
actcctccca	aatcccctct	acccactcaa	attcttatct	tgtattcttt	atcattatta	54900
tacatatgtg	tatatatgtg	tgtgtgtata	tatatatata	ctatatactg	ctaattagta	54960
acatttagtg	ttattcattg	ttgcatgttt	tcaatgtgct	ttccaggagg	ctggggggat	55020
ggctcagtg	gcaaaattct	agctgcacaa	gcctaaggac	cagggttcag	atcccccaata	55080
taaaggctgg	ctggacatgg	tggcttgctc	atgatactag	catgcttgct	ggaagcaaag	55140
acagggaaatc	cctgggagact	tagaatctca	gaagtgatct	gggctggaca	gactagctga	55200
actggccagc	ctgggttcca	tcaagaaacc	ctacctccat	aacataaagt	gtgatggaga	55260
aaggcaccta	atgtcaacct	caaaccccta	ctctcatgtg	cacacacata	catccacacc	55320
acacacacac	acacacacac	acacacacac	cacacacaca	cacacacaaa	taaataagta	55380
aataaataaa	atatttagct	ctccagacca	aatcttggtg	aaacccatgc	atttgcattt	55440
gtgtgtgtcc	tacaaacact	gaaggttaag	aagcatgctc	cttagtaatt	ttatagcagt	55500
ttgcgtttcc	agattgaaaa	cagattctat	aggctacaca	gtgctaaatg	gattatgctc	55560
agatacagat	tgaagagat	acagattgaa	aagggtcggg	gtctgggcca	ggatgacggg	55620
ccaactgac	tttgccgggg	cttgtccttc	aggggaagggt	tacaggattc	accactgggg	55680
tgtggcctat	ctgctgttag	gacctgaatt	gcctggagtg	tttctagttc	ccactagttg	55740

ttgaacttta	ccttgaacct	ctgctcccag	ggaagtcate	aggactctgc	catccctgga	55800
gtctctgcag	aggttggttg	accaacagct	ctccccaggc	cttcgcccac	gacctcaggt	55860
aaggggtttg	gatttggaag	gatgcaattg	ctataggagg	gactctgaag	gcagacagac	55920
gcaccgcctc	ctcacgttgg	ctagtctaata	ataaacatcg	cgggtggatgg	tgaggataga	55980
ctccatgccc	ttttgtgaag	gcatttcctg	gcacagctc	ctgacttcag	acagtttcac	56040
ccatcagaca	aattgcctgg	tggttgaggga	ggaggtgagc	agggccattc	ccatcatctc	56100
tcctcagaaa	tggaaggca	aggaaaacat	gaggttcttc	agacacttaa	tccttgggac	56160
tgcaaatagg	tggtgcccct	cctccacagc	tgctcacggc	ggggcaggag	atgagggcca	56220
aatgaagcat	agatctagct	atTTTTTTTT	tagtgccctc	agtaaattta	aaatcaaata	56280
aggaaggga	ccatagatctt	tatgttatgg	cattgtttaa	agtgaagact	tgtagccagg	56340
gtgtggtggc	gcacaccttt	aatcccagca	cttgggaggc	agaggcaggc	ggatttctga	56400
gtttgaggcc	agcctggtct	acagagttag	ttccaggaca	gccagggtta	cacagggtttc	56460
tctatcacaga	gaaacctgtc	tcgaaaaacc	aagaaaaaaa	agaaagaaaa	agaaagaaag	56520
aaagaaagaa	agaaagaaag	gaaggaaagga	aggaagggaag	gaagaaagag	gacaacatgg	56580
tctaggggtc	agagagcaga	atctccaaaa	acaccaacaa	tgccctgctgt	aaatgtatgt	56640
cgttgatttg	gggatgttgg	cctccagctc	accatttccct	gccttagcct	ccaaagtgtc	56700
aggattatag	ccttgagcca	acacatctgg	cttacgccta	ttgtgtgtgg	aaggggagtg	56760
ctgagtgtgc	tcctgtgttt	ggtacttata	tatgaatata	tgatatatac	catgtacgca	56820
tacttgcatg	tgaaggccag	aggccaatgt	cagctgctct	catcttatcc	tttttattac	56880
attgtattta	tttgtttgtt	tggtttgttt	tttcatcgta	cgcagtcagc	cactcatgag	56940
catgtaacag	cacaggtagt	aaggtagact	tgaggaggtc	agttctctcc	ttctgtcact	57000
tgagttccag	gaccacactc	cagccccag	gcctgggctg	taagagagcc	atcttactgg	57060
tcctctactt	tgtcttctga	gatagcatct	agactcacgg	aacctggagc	tcctctagat	57120
ttacattggc	tgccagctg	atgcatttta	aggtaaaatc	ttcattccat	ccctacccca	57180
cttccactcc	cagtgtctga	gttcgggaca	cctgccacca	agcccagttt	ttcctggatg	57240
cagaagctcc	aaactcaggt	tcctcatgtt	gcattggcagg	cacattttca	gttaagcctt	57300
ccccccagct	cctttaccct	ggtctctgaa	tggggggggag	gctataaatc	aggctgctct	57360
cagacattag	gtaggaaata	gacctatata	atgaggaaag	atattcacct	gccccatggg	57420
taccaggaag	tgatgtccaa	ctcctctttt	gcttatcagg	agaaatgctg	actactacct	57480
ctggtaattt	tgatgttggg	aggaacaggg	acattcatag	gaccccatc	ctcgtggttg	57540
agagtggaga	cagggtttct	gaagggcagg	agatctgtgt	agaaaagatg	gatgtctgtt	57600
tctgaaggga	aatggaggta	gagtcgacct	ggaggtgggg	ggaggtgggg	gatggttggg	57660
aggaatgaaa	ggaagagaga	cggcagttgg	gatgtattgt	ataagagaag	aatcagaaag	57720
aaaaaaagaa	aagctacctg	cacccttcaa	gtgttctctt	gtgtgggagg	ctgtctcagg	57780
gactacatgg	gcaccgagag	gcacagtgga	gggtaggtag	ttgatgttgt	gtccctgaaa	57840
acaaggacag	gaaatctgct	gcattggccta	agatggcaaa	atgtggcaca	atcaagtaag	57900
gcccaggatt	ctgtctgtgg	tgacagacct	ctgtagaatg	agctcccagc	attcccactt	57960
gctgtgtgga	gacagcatgt	tgacagacct	tgtaggagtg	agggtccagg	ccaggaggat	58020
gtcaaccac	caccatgtag	ccagtgggct	ggggggagctt	ggggccacca	aggagcttga	58080
gcagactgac	agtgggttat	gtacacaagt	gggcgtgtca	cacaaccgtg	caacacagag	58140
aaaatccctg	tgatgacaac	ttctaaacca	ccctgaggca	aaaggagtag	acaggggatt	58200
agagcctagc	atattggagt	cgagtgggca	tgacagctctt	ggaagcgtga	ggaaggaaat	58260
ttcctggaag	gataggttgt	cttctctagca	gcctcgtcaa	tagatgtcaa	tgatgtagggt	58320
agtacctgct	acaatcctgc	ttcttcagaa	gactgaggca	gggggattac	ttgaaccag	58380
aagttctagg	ccagtatgga	caacatagca	agagttctgat	ttaaaaaaa	aaaaaagtaa	58440
agagggaaac	caaatagggt	acgtgccaca	ctagtgtctt	cctgctccaa	gggtcctggg	58500
cacatgagct	tgcttagtgc	cagaaaagtc	agaggaggag	agggcagaca	gagaccctcg	58560
tctccacctc	ctttgactga	ctaattggggc	tggaatataat	ctgttttaca	aaaggacagc	58620
ttttcagagc	tggttctatc	taaggttgct	ctgaatagcc	atctcgaaat	atgccagaga	58680
agaatattta	gaggcggcct	attttgggtc	cccacaaaaga	tttcacaggg	aaaatgtatt	58740
tgtgttctat	ttatacaact	aaaaatatgc	atcagcccgg	ggaaactggc	tctttgctgc	58800
ctttgaagtg	aaggatgggt	taatttctaa	gaaagtaaaa	gcaaatgtag	tgagggcacc	58860
acggatgtcg	ccagacacca	gcgtttaagt	ggcttgaatg	gaagagcaca	ccccaggtat	58920
ttgaagttag	tgaggagaga	gaccaggtct	ccagagttgg	gcctgcagtg	gccagggtaa	58980
gtccaagggg	cagtatgacg	cagaacagag	tggaacacct	taccagtagt	agaattcagg	59040
tctcattcct	aacctcccat	aaagcagaga	atattgcact	ctctttctct	ctctctctct	59100
ctctctctct	ctctctctct	ctctctctct	ctctctaact	cacagagatc	ctcctgcctc	59160
tgcttcctca	ttgctggaaa	taaaagcatg	tgccaccaca	tgacagctct	cttgtttgga	59220
gagagaaagg	cagacagaca	gacagacaga	cagacagaca	gagtataatg	tgtagcagatg	59280
tcacacagaac	tccgaagagg	gtgttggaac	cccagaaacc	ggagttctag	gcagttgtaa	59340

gctgtccagt	gggtgttgcg	aactgaactc	aagtccctcg	gaaaactgga	aagtactctt	59400
aaccactaag	ctatctgtta	gtccccaaag	aatgtcttat	cttgataggc	cttcagatct	59460
cacagtccag	ggggtgacc	tgctacaagg	ggtagaggaa	agaaaactgg	ctccaggcct	59520
gagttccagag	tgacatctcc	agagttcttc	cttcccttgc	cgtgtagtaa	tttccctaccc	59580
tacacgagga	gaaaaggaac	agatatgtcc	agtatgcctg	gcatcttgaa	agggcactga	59640
ctcttgccgc	tgtagagccc	ctttccttgg	ggagtgcaga	gaagtgctgc	tagagaggtt	59700
caaaagaaac	acaacagtca	aaatagttgc	tggccaaggg	agggcatgtg	ccctgtatcg	59760
ggcctgcaaa	gccctttcaa	cagtgtagcc	aaggccatgt	ttgacagtac	aagcctgtaa	59820
gcccattgcat	agcaagggc	tgggataaag	ggctactggt	caagttagtt	atatacacat	59880
caagtttggt	catcttacat	ccttaggtaa	gggtgggtag	cttgtagct	ccactctcc	59940
agcaggaaag	catgtccacg	gtaaggtaga	tactgtagag	tttagctttg	ttttgacctt	60000
tctttcttcg	ttgcttggga	ttgtacttgg	gaccttcaact	catgcagccc	atgtttcagc	60060
cttgcataaa	tctataaata	tatgggttga	gtcagggtcca	gcatatgtgg	gtaccagtcc	60120
cagtctggta	tagcaattta	ttagctattt	gtcttttagac	aagtaactat	agggttctga	60180
gctgtaatgg	ccaccaaaaca	ggctccttta	gtatctttat	tggccaaaca	ccagggttaa	60240
aacactgaac	acagtatggt	ggtttcaact	tgatcggagt	ccacattcta	gcagggggaat	60300
tttacacaca	actcacaat	gatacagacc	acagcagggt	gatttgacag	agacaacgaa	60360
gggcttgccct	gagctgcagt	agtccgggag	cagctcttga	tgggatctga	gatctagatg	60420
gcagagcgcc	cttcctagat	gaaatccaag	cctctgggtta	ctcctcagca	agagtagctg	60480
gaggaagacc	agaagtgagg	aggaaaaggc	tcagagtggga	gcatctaagc	ctgcagcacc	60540
tccagagaca	ggctttggag	cttgccctgtg	gttcagccta	ctgtgggaag	agaccagtga	60600
gggttttaag	ctgtggggat	gggtatatatt	tgggaaggcga	acaagaacac	cagcagctca	60660
ggcaatgtag	gagggcggga	aggtgatgga	tgatcgagga	cggccccaga	gattttaagg	60720
actgtgtagg	tgggggtgaga	agcaactgtg	gtgaggtgga	gttagagggga	agaagtgaag	60780
atctattttt	ttggccatga	ggaatttggg	gtgcacacac	acacatacat	acacacccca	60840
tcatgtcctt	gtctacagac	ggggtgataa	tgggtactgga	aggtagaagg	tggaaagagg	60900
accagcaagg	aaggtatcca	gtgcccaccc	cacagctcac	cctcagacag	acctgttct	60960
cattttcatt	acccaggtgc	ccggagaggc	cagtcctatc	accatggttg	ccaaactcag	61020
ccaattgaca	agtctgctgt	cttcatttga	agataaggta	ctcacagggtg	tcctgtagca	61080
cctcctgatt	tctctgcagt	cccctatttg	cactagtggt	ccaagtccca	ggttctgcatt	61140
tcctatgact	cgtgacatt	cctgctgtcc	acaggccaag	tccttgctgc	acgagggtc	61200
agaatctacc	aacaggcggt	cccttatccc	tcgggtcacc	tttgagggtga	gtgattctgg	61260
agcagtattt	ggcgcgtcgt	tcctgtgaag	agagtgcctt	tgatttaggc	tgtgaggagc	61320
gaggaaactca	atggcagaga	atagtagtca	ggccagccac	ctcgctgggg	cagtgggtaa	61380
acactaaaat	cgctgcctaa	atgcaccagt	gaagtacccc	tcattgggag	ctgcaacgaa	61440
agagctcccc	attggtccaa	gatggtatga	acctccttaa	aataggaagg	agaagtgaca	61500
tcaaactgag	ggccacacgt	tgccataaaa	atggtgtctg	gtacaggaa	agtgggagaa	61560
aacaattatt	tttgttttgg	gggtttgggt	ttttgtttt	gtttgtttt	gtttttttgt	61620
ttttttttt	tgagacaggg	tttctctgtg	tagccctggc	tgtcctggaa	ctcactttgt	61680
agaccagggt	ggcctcagac	tcagaaatcc	acctgtctct	gtctcccaag	tgtgtggatt	61740
aaaggcgtgc	gccaccacca	cccagctttt	tagttttttg	tttcttgttt	tttgggggag	61800
aaaacaattc	ttatgatatt	cttaaaaatg	actggcttaa	gtggagacct	agaaagggtg	61860
ttctttctct	ttctttttct	ttctttcttc	ctttctttct	ttcttcttct	cttcttctct	61920
ttttttttt	ttttttttgt	ttttcgagac	agggtttctc	tgtatagctc	tggctgtcct	61980
ggaactcact	ttgtagacca	ggctggcctc	gaactcagaa	atccgcctgc	ctctgcctcc	62040
caagtgtctg	gattaaaggc	gtgcgccacc	acgccagct	ccttctttct	ttttttaaag	62100
gtttatttat	tattatacat	aagtacactg	tagttgactt	cagatgaacc	agaagagggc	62160
gccagatctc	attacagatg	gttgtgagcc	accatgtggt	tgctgggatt	tgaactcagg	62220
acctctggaa	gagcagtcag	tggctcttac	ccactgagcc	atctcgccag	cccagggtga	62280
ttttttctaa	tggaccccca	tgattcttgt	acttccagct	caacttgggt	ctgcctttat	62340
gagacacccc	actgccacc	ccaaagccct	cttcatccct	gagaggagtc	tcacaaaaaa	62400
cagatttcca	ttctctctta	ccaccttgggt	tttttaaat	tcacatttat	ctatcatctg	62460
tctgtctgtc	tgtctgtctg	tctgtgtgtc	tgtctgtcta	tcgtgtgtgt	gctgagttag	62520
tgagttagtg	agagagagag	agagaatatg	aaggaaaagg	acagcttgca	agagtggatc	62580
tctcctaccg	tgtgggtccc	aagagattga	ctccagattg	ccatgcttgg	tggcagattc	62640
cttttcccac	aaaaccatct	tgctgacttt	ctctactgct	cctggcttta	ggggggcttt	62700
tttgttttgt	tttatttttg	ttttcatgag	ataggacccc	caaatagcct	ccttgagtct	62760
acctaaacct	tccaccacac	ccagcttagc	ttttcatttt	ctaatacatgt	ctgcgtagga	62820
atcgtcttcc	taattattac	cagtcogatg	tcaattagcg	gttgtaacaa	ggccccagag	62880
tgtgaatcta	ttgctaagat	tctttccgga	agtagatgac	acatcagatt	aataagtggt	62940

gtcaaaactca	tcagtgaatg	aaccattttcc	caacagttca	tctgtttcct	ccaccaaatc	63000
tggttggtgga	gcagctgggt	tacacatggg	cgagctgggc	tcagccgcgg	tgccagccat	63060
ttttccagca	tgctatggaa	tataggggtg	ttatcagggc	aggggtggga	aagcctggga	63120
atagcgagcc	agatagcctc	acccgggaca	ttgaccatcc	cagccatgta	gataaatcag	63180
agttgacatc	accaaagcca	gccactgact	gcatgtctta	tgctctgaag	gtgaagtcag	63240
agtccttggg	cattcctcag	aaaatgcac	tcaaagtggg	cgttgagtct	gggaaactga	63300
tcgttaagaa	gtccaaggat	ggttctgagg	acaagttcta	cagccacaaa	aaaagtaagc	63360
caccaccccc	aaccctgcaa	cacacacccc	atgttctggg	tcctgtgtga	gagctcttat	63420
gaatgaccag	tgactagatg	tcaacagcga	ctgtcttagt	tacttttcta	ttgctatgac	63480
aaaacaccat	gaccaagacc	acttacaaaa	taaagtgttt	gatgtgggac	tcacagctcc	63540
aggggggttag	agcccatgac	catcacatg	gggagcatgg	taccaggcag	gggagtgtgg	63600
cgttgatca	gtagctgaga	aactacgtct	tggttcacaa	gcataaagca	gagagggcaa	63660
actagaaatg	cogtcagctt	ctgaaatcac	aaagccccgc	ccccagtgcc	acatctcccc	63720
caacaaggcc	acacctccca	atccttccta	aacagttcca	gcagttggaa	accaaactt	63780
caaatatatg	tgtctatggg	agccattctc	attcaaacca	cctcagtgat	tgtcagcaaa	63840
agccagtcct	accccttttg	cagggtagtc	cctgtccatg	ggtgctcccc	aagccgctcc	63900
ctgccttagg	aagtcttttc	ctgatacaag	ggcactcctt	aggggcctcc	cctgcatact	63960
tagccgcccag	ccctgaaaat	gagattataa	atttgccaca	actgccccat	agcagtcctg	64020
tgtacacacg	gtgatttcat	gtgtctgtgt	gactgtcctc	ttcctgcaa	gcatgcagtt	64080
acctcactgt	gccgggacag	gccctgtact	cacgggtattg	ttggcagagc	tgctgggtta	64140
ccacccatag	caaatacacta	accctgtgga	gctgaagtat	atgaagcctg	cgtctaaatg	64200
gtccaaactc	tcactgccgc	tctgtcactt	ggcagctggg	ggaccttgac	ccgtctctct	64260
gaacctcagt	ttccagttct	cagggtatgag	aaagattgag	tgaacaacat	gagggacagc	64320
ccaacaccca	gcttcccttc	caccagctct	cctcatggac	accgttgagc	cactgcagct	64380
cttctgtctg	ttgtactgtc	cagtatttgt	acacacagca	ggaaagtccc	ccatgacagc	64440
tggcggagaa	gctgctgtgg	gcagagcagc	atccagagaa	gcaagccacc	catatggtcc	64500
ctgccacat	gcaatttgta	ttctgggtgca	ttgaggggtga	ataattctca	gctaaatcta	64560
ccaatacagt	ggcttcagct	actcgggaagc	tttgaagaaa	tcaggacagc	atgataggta	64620
aagatccttt	gggaagagag	aatggacaga	ggaggctagc	ccttttaagc	aatgagcttg	64680
tgagctgaga	tctcaaaaaa	gctactgggt	ggacaggtgc	ctgaggggaag	agtctcccag	64740
cgagtgcata	aggtggagtg	acaggatcct	aaagaacaga	aatagatgtg	gggacagtg	64800
agaatggaa	aggagggcaa	tgggaattgt	acctcctga	acctcctgaa	gaggactttg	64860
gaaaaactcc	caagcaagaa	gttttaggct	agaccttggc	gctgttgaa	tatggatgca	64920
ggtggttctg	agggtagccc	tgggtgaatg	ctcatctcca	tatgggggtg	gacaagccca	64980
ccttcaaat	caggttctac	ctcctgccag	ctgtgtaacc	tagcctctct	gcgcctccag	65040
tctccatctg	taaaatgggg	atgaagacaa	cagatgtcag	gtggagattg	catgcttatg	65100
agagcaactt	gcacatctcc	tcactcctct	aagttgagtg	agagaatgcc	caggcagggg	65160
ctagaagggg	gacctgttga	tcctacacct	cagcactggg	gaagaaaagg	gctctttttt	65220
tcctctctgc	attttgagtt	cgtggcagaa	gccaaggacag	attacccaaa	gaaaggagaa	65280
tggattgatc	acatgtaagt	ctcgtgacgt	gggaagccct	acagggaag	caggaccagg	65340
aggatggcta	agcagagtgc	tgtagccaga	tttccaaaag	gtgggaagct	agaaagtgat	65400
agcagagagg	ggctgggggc	taagtttagg	gcaatgggaa	gagcttggcc	aggcccgtaa	65460
gctcagactc	ctctcagctc	tgtttgtcct	tggcaataaa	tgccctccct	tcgggtacag	65520
agagcatact	tgtcacatga	gggttttatc	tccagcttag	aggaagacca	gaaaatcctt	65580
tctcctcctc	catccttctt	gtcaggggtgc	tatatatttg	ggtactgggt	cctcagctcc	65640
atcagtaaat	agcacaggag	gtattagaag	tcaagaactg	agtttcccag	ctcagtcact	65700
gtgacttggg	ttggtgccc	ttctggtcct	tggttgcatg	ttgctacaag	gcttaaaaga	65760
atgaatatca	aaatgtgtga	tgcttgccgt	ggagagagtg	tttgggtgcaa	ttggccactg	65820
gcggaggagt	gtcatggta	caattatcct	tccatgtttc	tgtgaggaca	gaatggaatg	65880
ggggtggaag	gtaaagagag	gaagtaggaa	agacagcctc	tggactctgg	agtagcaagc	65940
tgtagggaaa	ttacagagct	gagcttacat	catcaggaca	gggctctctc	tgccctactg	66000
tgtctagttg	gggagtgaat	gatggtttga	gcgggaaggc	tatgccagaa	gccgttggct	66060
tagaatagt	catactgctt	gtacagggca	aataagttca	gagattcatt	gcagcagggc	66120
cactggtgac	agctacttag	agcaaaactga	actcccaga	agctctgaga	gctaggggtt	66180
ggtgctagct	gggtgctctg	gctactggga	tggtgtgtgt	atgcactgcc	cgtgctacat	66240
agatttacc	attccactgc	gtgcccactgc	ttcagagcct	cctgtgtgcc	atgggtgttt	66300
tgtacagctt	tctctgtcaa	attaaaataa	attaataata	aaggaacaat	gttgagaggg	66360
gagagagaga	gagagaggat	gtggatagat	atgttggggt	ttttttctg	ttgggtgttt	66420
tgtttcaaga	cagaggctca	tgtttcctat	gctaggggtg	aattcgctct	gcagctaagg	66480
tgagtcttga	cgcccatgga	tcctgggttaa	ctgcattcct	tacataatct	agatctcatt	66540

ggtttttagga	aagcaggaaa	tacaaatgag	cacatgttaa	ataaaatcat	atgttgtact	66600
gtactctccc	agacagcctt	gtactgtact	ctcccagatg	ctttttaaag	taacaatgct	66660
aggccacatg	tggtgttaca	gccttttaat	cctagcacac	agaaggcaga	acaggctggt	66720
ctctgtgagt	tcaaggccag	tgtgatctac	atagttagcc	ccaggccagc	cagtgtgca	66780
taataaagac	cctgtcttaa	agaaacgggg	aggaagggga	ggtgggagag	gagggaggga	66840
gggaagaagg	aaaaagaaaa	atagtacaaa	ctgagtgtta	tggtacatgg	ctggaatcat	66900
aacaccaagg	aattttgcac	ttttcttttt	ataaattttac	attgtgtgtg	tgtgtgtgta	66960
tgtgtgcaca	tgtgtgtatg	cacttgccctg	tgtgtgtctca	cccatgtaaa	catacatgca	67020
aatgaccaga	agaggctgtg	gatcccctag	agctggaatt	acagttgtaa	gctgcttgac	67080
atgggagctg	gcaaccaaaac	tggtgtcctc	tacaagagat	acatactctc	ctaaccactg	67140
agcctcccag	cccacactcc	aagttaggggt	acacagtgtt	tcttcttgtg	tcacaggaaa	67200
ggaccaatgc	tgagtgtttc	agaccacaa	aagccagaaa	tgtcatgatt	cagcctcata	67260
gttgcaaate	taagctgtcc	tctcttcagc	tatgtatgtg	tgtgtgtatg	tatgtatgta	67320
tatgtggctt	ttttatgaaa	cctatcttta	attggaaaag	tctcatcttt	gtgtctttat	67380
tcttatgagc	ttgcacccat	gaaggaaagag	tgggaccagg	caaaaatgag	tgagaagaca	67440
aaacacacga	tgggatgagt	tacacagata	aagtgaata	aggacatgag	cctatgaaat	67500
aagctggaga	gacggcccg	cagctaagg	cactgggtgt	tctttcagaa	gaccaccgtt	67560
tggttcccc	cctcaagaca	tgacagcttc	cagccatctg	taaccctagt	tctattggat	67620
cacatgccct	cctctgacct	ccatgggctc	tgcgtgcatt	cagtgcacag	atatgtgcag	67680
gcaaacagct	gtacacataa	aataaagatt	tgaatcaaag	ccagaaatcc	ctttaatccc	67740
agctcttggg	agacagaggc	agggggatct	ctgtgagttt	gaggacaggc	tagtctatat	67800
agagaattca	aagacaggaa	tatgtagaga	gatgtatgta	tgtagtctca	aacaaaacaa	67860
aatagaattt	atttttaaga	tttattatat	tgaagtacac	tgttaactgtc	ttcagacaca	67920
acagaagagg	gcatcagatc	tcattacaga	tggttgtgag	ccaccatgtg	gttgctggga	67980
tttgaactca	ggacctcg	taagagcagt	cagtgtcttt	aaccactgag	ccacctctcc	68040
agccccaaaa	tagagttttt	aattttaa	atttttgaat	taataggaag	agggagaaga	68100
aaaaaaaaag	agttctgtac	aagggttaaa	ggcagggcat	caggaattca	agccagccgt	68160
ggctacatac	aacctgtttt	caaaaatcca	aaaagaaata	aacagatttt	caaagtaaaa	68220
attttaaagg	actggagata	catctctgtt	gtccagtgtc	tatcaatcat	atggggccc	68280
cttgtagtgc	cgaaataaaa	taaaaacaac	cagaaattaa	aaacttgagg	gaaggagct	68340
cttagaaggt	gggtttgggt	tgggtttttg	agccagagtc	taatgctacc	tcggttgcct	68400
ggcagctttc	ccacaagagt	catatggagg	tactcaatgc	ccccacctct	gtcagtcag	68460
acaagtctac	cagggacttt	gaggaggggtg	ctcaaagtgg	ttattttgtt	gttttagctaa	68520
ttagttttgt	aatcacagta	atcaaacc	aggacctatg	catgtcaggc	aagtggtagc	68580
caaggaaactc	actcctcatc	ctcctgcctc	tacctcccaa	gtgctgggtta	tggcgtctgc	68640
cagtaggcct	agctccagg	aagggtctct	caaaaacaaa	caaaaacact	ccaatagcaa	68700
aaaatgacaa	atgggatttt	ataaaattaa	aaaactctga	ccagcaagg	aaacagttag	68760
cagagtgaag	agacagaatg	ggagaaagtc	tttgccagtc	acacactgac	aggggatacaa	68820
tgccagtgat	ggcgccacac	ccgcgtttat	cagagcgaca	tttgttatag	ccaagttatg	68880
gaactgacct	tggtgcccaa	cattgaagga	agaaaatgtt	cctgaacatc	atgggaggag	68940
tattactcag	ccataaagaa	caacacgata	ctattttacca	aaaatggatg	ggactggggg	69000
gtgaggggggt	acaattagg	aagaaaggga	ccaggacaga	aacaaagagt	aactgtggaa	69060
tgatcaaaga	gtgaatgaag	aggttacaat	gaatagatat	gttaatatat	gctcatttaa	69120
aagtcaaaca	tgttttagaa	aataattctt	ggaaagctac	tagagatgtt	accagaatga	69180
aggtgtaaac	aaagagagag	ggagacagaa	tgagagagag	aagggtccaa	ggagagggac	69240
aaaggagtag	atgccagggt	gctgaaggag	gaagggttcc	aacaggcctc	aaagtcagca	69300
gacctagagg	ctgtgggaaa	gcacttactc	ataggacaga	cagcaaagtg	cttagtcag	69360
tggggaatga	gaggggagtc	tgaatggagg	aaagcacata	gcaggggggt	aggagggtag	69420
cttagcatgc	agtgtgaccc	tccgtaccac	caaaaaagaa	aagaacatgg	tcattgtaca	69480
gttataaacc	caggaaaggc	caatataata	tgccttttct	catactgcc	atactcacc	69540
catactcctc	cctctaactc	ctctcagacc	cacctcacc	ccctccctcc	tcccaacttc	69600
ctgtcactc	actgagtc	tgtgatgggt	tcctatgtct	tgggccagg	agtggcacta	69660
ttaggagggtg	tggctttgtt	ggagtctgta	cctgtacca	tcacagtgg	tgtgggcttt	69720
aataacctac	tcttagctgc	ctggaagtga	gtattctgct	agcagccttc	agatgaagat	69780
gtaaaactca	ctccttctgc	accatgcctg	cctggctgct	gccatgttcc	tgccttgggt	69840
gataatagac	gtaacctctg	aacctgaaag	ctagcccag	ttaaatgttg	tccttataag	69900
agttgccttg	gtcatgggtg	ctgctcatag	cagtaaaacc	ctaactaaga	cagaaattgg	69960
tatgtgctcc	tgggtatgag	accatccct	gggacctggc	actacctgga	gctacatcct	70020
taagaaactt	gactttctct	ccccagcagt	catctgccaa	tagttctca	gctaggggag	70080
gggctcatgg	gccaactca	gggaaacgtt	tttaagttac	ataggaaaa	gaacaaccag	70140

gactgggtctg	tcaggagcct	tgtcgtgggtc	acagtgacat	aagcattgag	tgtttatgta	70200
gctaaaaatac	aacaataaac	caggtgggtca	tgcatgcctt	ttatttatgt	ttctattttt	70260
taagaaacat	gagataaagt	taaaagagtt	cattgaaaaa	cttgaaatta	gggctggaga	70320
gatggcacaa	cagttaagag	cactgactgc	tcttcacagag	gtcctgagtt	caattcccaa	70380
caaccacatg	gtggctcaca	accatccata	atgaggtctg	acaccctctt	ctgggtgtgtc	70440
tgacagctac	agtgtattca	tatacataaa	ataaatatat	cttttaaaaa	aaaaaccttg	70500
aaattagttg	ccttcagggg	gcaggacatt	ggagggagca	gagtggaaag	atactattgc	70560
caacaacaac	aaatcttaaa	gaataatttg	attcattttt	aattttatcc	ttattatttt	70620
taattatgta	tgagtgtatg	cctgggtatg	tgaggtgctc	caagagacca	aaggagtcgc	70680
ctggagctgg	agttccaggt	ggttgtagag	tgctgtatgt	aggtgctagg	aattgaactt	70740
ggatcctctg	aaagagcagt	acatgttcct	agctgctcag	cgcctctccc	agccctcaa	70800
tttttgatct	aaaaaaaaag	ttgaaagcta	caaattttct	gtgaattttt	cacaaattaa	70860
gcaagacaga	gaaaaggtag	ggaggtgaag	tggcatggaa	ggagaaggag	aaaaggatgg	70920
acagtctcat	ggtgtgtgga	gcaaagctaa	ctctagctgg	ttctgggtac	cacactttaa	70980
aatgagcccc	cagggtcttag	cccagggtatg	tggggggccag	attggggctt	gggtgcctcg	71040
ggagcaatgg	acagtggaca	tagcggataa	agaaacaaat	gcatgaatcc	ataaacttct	71100
gtttttctga	agttcctgag	ctcatatagt	cccagaagtt	tctaaacaag	ttggtgattt	71160
tggtggagac	ggagaaggag	aaaatcctga	ggaaggataa	tgtttttget	gactctaagg	71220
taagtgacaa	tagacatcta	atgtgaggtg	aaaaggatca	tgggccttga	ggaccgtgct	71280
gtgctaccag	ggctggcgat	cattccctcc	ctgggttcta	cagcaacctt	ttcattcctg	71340
ggaagctgaa	catagctagg	taagaaccag	gctctggggc	agctgggact	tgacaggttt	71400
tagttccagc	gtcccaggat	cacaagtact	atcttccagg	aacaaggaga	cactccttgt	71460
ctcaggccca	gtgctgagct	ggccaggatc	ctagccttta	ttaacagtgt	ctcttggcga	71520
gatttaggag	ccctttgggt	aagctctgga	tttaagacct	gaaccagtgc	tcccctgtca	71580
cttctgtgtg	acttctgttt	ttaaactctg	ccctgctctg	ctaaagtaac	ccctggatgt	71640
gcttcatcca	atccagatgt	gtggtctctg	catcttgatt	gctcacctca	gaagtcccct	71700
caaattgggt	gattaccaca	cagaggcctt	gctgttcacg	ctcagtgagg	cgctaaaagt	71760
ggatagaaga	acagatagtg	tgggtccaca	tgttgcccca	cctctggggg	gtgggacatg	71820
ccacttagga	aggtgaaaca	caggagtgtt	cctgtgctta	tcccataccta	ctgccatcag	71880
gtgctgggct	tcagggaaga	accaagacac	cattctgggg	gtaggaagat	gcaatctaag	71940
ataggggaga	ttgaagtgg	ggtttcctga	cttccgggcc	tccagttgct	tctggagaac	72000
tgcacagaga	agtcggtaat	gaagtgtcag	ggatgtcagg	gcgacacagc	gtgataagg	72060
ggaggggggg	ctgtgtgtgt	gtggctgagt	tgcaaggaa	tgagagcaga	gttcaggag	72120
ggaccacacag	gctgatgact	gggactagcc	cagggggcca	ggggaatggg	aactctgggt	72180
tagcttatca	acaaaggggt	gcagatggac	tttctatggg	agacttaagc	ttcggtctca	72240
tagatgaagg	ccttccccat	gctccccatg	gcagttctca	atgaaagaga	cagtccttgc	72300
ttccaaagtg	ttattgggtg	ttcatcatgg	aaaatggctg	cttaccact	tctctgggtg	72360
ggccggagat	taaatacctt	cttcaagaag	gaatgtcggg	tcaagaagg	acactttgtg	72420
gctaaactct	cagggcctct	aggtacctac	ctcattagta	tagagaattc	tgttttgggc	72480
tatgtgaaca	acatggggaa	tatggtatgt	gttccaggcc	aggaatctgg	caggagcag	72540
ggaggcacct	accatcttaa	gtgtgagttt	ctggagtctt	ctgttctcca	ttaccaagtt	72600
tccctggcatg	gtgaactgtc	ccacaagata	gggcacatag	cttagcatgt	gtggctgact	72660
cacgggtcag	acattggaga	caagctaagg	gaaggctccc	agtagagctt	catagaggta	72720
acccccacca	ggttcccaag	gcactcaact	tcactcatcc	tgggaggga	tatgaaggta	72780
cagtggcagg	catgaccga	cgggaacagt	agcaagagag	tacagaaagt	agagaatgtt	72840
cacatggtta	ggcagcatc	taggcttcag	gagcataggt	cccaaactgg	ctatttctgt	72900
cccattgctg	caagctcagt	ggggttgaga	cactgacctt	tgagccatcg	ctggttaaag	72960
gcaggtttca	ataaagaagc	tcttggttatg	tcagtgtgag	ctggtatctt	tagagtgggc	73020
cactgtgata	tttgcaatga	tgaggttgaa	tcatatgaaa	ttgctggctt	tgaaccagg	73080
tgagggacac	caggcatccg	gcctttcata	ttggcacttg	tcatagttag	ggttttactt	73140
ctgtgaacag	acacaatgac	caaggcaact	cttgtaagg	caacatttaa	ttggggctgg	73200
cttacagggt	cagaggttca	gtccattatc	atcaaggcag	gaacatggta	gtgtctaagc	73260
aggcagacat	ggtgcaggag	gagctgagag	ttctacacct	tcactggaag	gccactagga	73320
gaagactggt	tttcaggctg	ctagaatgaa	ggtcttaagg	ctcacactca	cagtgcaca	73380
cttctcccaa	taaggccaca	cctatcccaa	caaagccaca	cctccaaata	gtgctattcc	73440
ctgagtcaag	cattcatagg	actgcttgaa	attttgtgtg	ttctcattta	aagtcagcca	73500
ccctattgac	gacagccctc	tagactaaga	gagagacgag	tctcatgctc	cgctgtgagc	73560
ataggagaaa	tacatgtaag	agatcagggg	ggcccttttc	ctaactgcc	ctttggcctt	73620
tgcttagttt	tatttaaaaa	aattatttta	tttgagggtg	gatgcatgtg	cctcagtcca	73680
tgtgtggagg	tcaggaggaca	acctgcagga	gttacttccc	tctctacat	atgagtcoca	73740



aggattgaac	tccattgcat	agtgtatgtc	ttcactgcgg	agccattgtg	ctggccccgt	73800
ttctctacat	tttctctcac	caaataagta	attcagtcct	acagaaataa	aataaaaaacg	73860
ttgcaaccaa	aatagttaaa	agtttaaaag	aaaagcagca	cggccttcac	tgctaaccgg	73920
ttctggactt	ccccctccgg	tcggcttaag	ggattattgt	cgtcatcgtc	gtcatcgctg	73980
ttgtttgttt	ttatggtttg	ccttggtttg	ggctttcggtg	agacaggatt	tttgcaatat	74040
tgttctgtct	agcctgggtgc	tcactattta	ccacagattg	gccttgaact	cacggcagtc	74100
ttcttacctc	agccttccaa	gtgctggaat	tacatggcca	tgccccacac	tttttttttt	74160
taaagattta	tttattaatt	atatgtaagt	acattgtagc	tgtcttcaga	cactccagaa	74220
gagggagtc	gatctcggtta	cggatgggtg	tgagccacca	tgtgggtgct	gggatttgaa	74280
ctccggacct	ttggaagagc	agtcgggtgc	tcttaccac	tgagccatct	caccagccct	74340
tttttttttt	tttattgcaa	tggtgataca	cattccattt	tgccaagtgg	cattttataa	74400
aaattatgta	ttgtggccac	ctttctgtgg	ataaagtaat	tctacacat	aatgttaaaa	74460
ctctgtgtgt	gtgtgtatat	atatatacat	acatatattt	gtgagtcctt	tacctgtgaa	74520
catgtctcag	tgtctaatg	tatcttggtg	cagttcctca	ttttttgctt	caaatcataa	74580
ccttgacgtg	gaacagctgg	gtccaagaag	ccatatactc	atagctttga	atgcatctct	74640
ccgggtgggt	gcgtctctca	tttatacaca	cacacacaca	cacacacaca	cacacacaca	74700
cacacacaca	cacacacaca	aagaccttcc	agtttttttt	taagctgacc	acatctttgt	74760
agaggccagt	gtgcaggag	gtggagggtc	tgctaagtgt	gtctgtcctt	cctccctcag	74820
aaaagagaag	gcttctgtca	actcctgcag	catagaaga	acaagcattc	ggagcagcca	74880
gagcctgaca	tgatcaccat	cttcattggc	acttggaaaca	tggttggtgc	cctgtgcccc	74940
ctccatccta	ccagctctac	ttgggtccac	cttcctgcct	cagcttctac	aagtggcaca	75000
agggggcacc	tctatctctc	agccatagtc	ctggtctcaa	tcccctctat	acaacctctt	75060
ttctagagaa	ttctaccttt	gtcagaggat	gatgaacaag	taaaataggg	tttaggatta	75120
gtgcacccca	aaagtagcca	aagtagctta	ctctgcagtt	taccaagggg	cccaggcct	75180
ggaataaagt	agctgggtcc	ttgcttatct	agggaggaag	tgaggggagg	gatagagaca	75240
tcagtggatg	agtagatgga	tggatgggta	agtgatttga	tggatagatc	ctagtggccc	75300
taagttaaag	tgccatataa	ataatatata	agggtttttt	taaagattta	tttattatat	75360
gtaagtacac	tgtagctgtc	ttcatacaca	ccagaagagg	gaggcagatc	ttgttatgga	75420
tggttgtag	ccaccatgtg	gttgctggga	tttgaactct	ggaccttcgg	aagagcagtc	75480
gggtgctctt	accaatgagc	catctcgcca	gccccaatata	taagttttta	gcataatgt	75540
gtagttaaaa	gtaaatgaac	aaaatatgta	tgctaaatat	aatggtagca	taatcatatg	75600
gtaaatatgt	aagtgaatat	atcatttggt	ataagtactc	aacaataaaa	gctaaggtag	75660
aaaatatata	ttggattaaa	tatatatcat	atataacaat	taccacatat	tattatacta	75720
aatatatata	atagtacatt	ataactaattg	tcaattatac	taaatgtttt	gggatataa	75780
ttgagctaga	attaaatttt	taaaatgtcc	agttaaatct	gaatctcaga	caaccattgt	75840
atggggcata	cttactctaa	aaatcaccca	gccaccagaa	attcaaattt	aatggatatc	75900
ttagtctatc	attaactcct	caggcactca	ccctcaggtg	ttcaggcatg	tggcacaagg	75960
acctttctag	ttctcttgct	cgtctgtgaa	gagttctttt	gggtgaacat	ggcgggaggt	76020
catcccatgt	ctgcccctcc	agcagcaaac	acatgggggc	gcccttccag	gagaagtcc	76080
atggaggcac	taggcaggag	ccctgctgcg	catggcaggt	ctccttatga	gtttcactta	76140
gagccccccg	gggggggggg	gggagtttgg	cagagatcat	tatgtggatt	attgcttct	76200
tgctctctctg	ttggctgagg	gacatagata	gatctcctca	gggtgtgggg	agaatatcta	76260
cccatactta	gggtagctcc	agaagtagat	taccatcat	tccttgaggg	agctgcttcc	76320
actcattttt	caaaacggaa	gctgagacaa	gttgacgccc	cctcctgttt	ctctgttgag	76380
cacctcttca	tctcagttct	tagctgaact	ataactttcc	tccaggaagc	ccttggtgtt	76440
tcctggctgc	acggtagaag	cctccatccc	ggacatgtca	ctcaccctcg	aaaatgtaaa	76500
ctcctgagtg	acaaacgggg	ccagatttca	tgtctcactg	ccaccctaga	gcctaggaca	76560
tgtctggcaa	cctgttcagt	tagatcaggg	ctccagagag	gatctcccag	gctttccatc	76620
ctttgaaatg	taatgttatc	tttacaacaa	gatctcactg	agctctgccc	atagctaaca	76680
tgtaagggtg	tatcagcact	gagtcctcca	tctagggcac	ttgcccccat	gatttttgta	76740
actatgtgat	tctcctcttc	ccctgttttg	aatcccccta	atgcccttga	cctgacctgt	76800
ctttgactca	ttcgggtgtg	ggttttgtctg	ttgaaggtaa	tgcacccctt	cccaagaaga	76860
tcacgtctctg	gtttctctcc	aaggggcagg	gaaagacacg	ggacgactct	gctgactaca	76920
tcccccatga	catctatgtg	attggcacc	aggaggactc	ccttgagag	aaggagtggc	76980
tgagactact	caggcactcc	ctgcaagaag	tcaccagcat	gacatttaaa	acagtgaaca	77040
gctggccagg	cctgggggtg	gaagacagca	gactctttca	agcatttcag	aagtcagaca	77100
ggatacttcc	aaagatgtat	aggattgtct	aggggtaccc	cactttcaga	gccacagatg	77160
tgcatgtagg	tggcaccctt	acaagttgat	agggctctga	gtccgcatc	ttccctactc	77220
ctgcttaaaa	gaataatatc	gccgggcgtg	gtgggtgcacg	cctttaatcc	cagcactcgg	77280
gaggcagagg	caggcggatt	tctgagtttg	aggccagcct	ggtctacaaa	gtgaattcca	77340



gggcagccag	ggctatacag	agaaaaaacc	aaaaagaaaa	gaaaaagaat	aatattttact	77400
tcctagatgc	attttcagtc	ccagttctca	tctctgaggt	gctttgtctc	atttctaggc	77460
attgtttgaa	gttccccctg	aaagctagga	aatacagaca	gggtgtctta	ctcccagggt	77520
ggaaaccgggg	tgcattgttc	gggttccaaa	ggtgtacctc	agtcctgtgt	tgcaacacca	77580
tctcccacca	ccaccaccag	gttgccatcc	acaccctctg	gaacattcgc	atagtgggtc	77640
ttgccaagcc	agagcatgag	aatcggatca	gccatatctg	cactgacaac	gtgaagacag	77700
gcacgcgcaa	cacctgggtg	gagcatagag	ggaaagccat	tcctgtgcat	gctcctcctc	77760
ctcctccttt	tggcaatgtc	ccaggataaa	ctgtgagagt	cctgtcctgg	gatcccttcg	77820
tctctagcaa	atccagaagg	ttttccttgc	aaaaactcat	ccagggtcta	taccagctct	77880
ggcaatcttg	ctaaaatgtg	ggttctgtct	cagtaagtct	cagatgggtt	ccattctgac	77940
aagctttccag	gtgacatagt	gccagccagt	ccagggatcc	cacttgagga	agtgtgtgta	78000
cttgtgtgtg	tcctgtatct	gggtgtgtga	tatgtgggtg	tttatgtgct	tctgtgtctg	78060
tgtatgtcnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	78120
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	78180
atgtgcgttt	gtgtctgtgt	atgtctctct	gtgtgtgtct	atgtgggtgt	acatgtacat	78240
gtgtgtatct	gggtgtgaca	tttttgatcc	cacaggagtg	gatcctgagt	agatagtcac	78300
tcactcagtt	gggggaggag	gcacctccca	aatcccaagg	ctccaggctg	ctccaccctc	78360
cctgttccct	ctctgtcag	tatcatggcc	cccaacatcc	ccatagccag	aaacagatga	78420
tagcattctc	cttcccttct	cctatgtaga	aaatcacaa	gctgtacca	cctgccagtc	78480
actggcatgt	cccacccccg	ctgacctatc	cactctcctc	tcttcccttg	gtctgtctct	78540
ctttctcctt	gctaaccagca	ttagtacctt	ttacctcacc	tcagcctctg	acccctcggc	78600
aagcctgcct	gcctgcctga	cattctgtct	cttctcttct	ttcctccctc	aggaaacaag	78660
ggagcagtg	gagtgctctt	catgttcaat	ggaacctcct	tggggttcgt	caacagccac	78720
ttgacttctg	gaagtgaaaa	aaagctcagg	taatgggagc	cattccctcc	atgcacccag	78780
acagccttag	cctgcacac	cctgatgctt	gccagcctc	aggaccttct	ggacatctgc	78840
cgtctgagca	aagacaactg	tagtacagta	gagacctcag	gtatatgtga	ctctgtctct	78900
gaggacagaa	aacatgttga	tgtcattatc	aacctagtga	acccctttga	gctgtgttca	78960
tttctaaagg	gaaataccac	tcaggagcat	cctatagctc	aaagctcact	ttgtctgcat	79020
gtgcctctgt	catgaccata	accagctatt	tcaattgctg	atctgtttcc	taatgggtag	79080
attcctgcat	ggaatctaga	ctctgtgttc	ccagttgggt	ttccagaca	cactcagggt	79140
accactggga	cagagcttga	gtctaaagtct	ctgtttccca	cctgttgctc	ctgacttccct	79200
actgtcagat	acaggacagt	acccatgcta	gatgtgctag	atggcccttg	ttttcttgaa	79260
ttaggagaaa	tcaaaaactat	atgaacatcc	tcgcggttct	ggccctggga	gacaagaagc	79320
taagccatt	taacatcacc	caccgcttca	cccacctctt	ctggcttggg	gactcgaact	79380
accgcgtgga	gctgccact	tgggttaagga	gactccacct	tgggtcgagc	cataggggac	79440
agaggctctc	aggagtcggc	tgtggacaga	aaatcaaaga	gacatgtgaa	cactcagaac	79500
caggctaagg	tagcagctct	tgcagagatc	tcaatacagc	tcctgtcatg	gatcctctgc	79560
ctttcttcat	agtcccttct	tgttgtgtcc	tcctttgccc	caggacttct	atggetgccc	79620
cttgtcactg	tcactgtctt	tccagtcagc	cagccagcca	gcagcagctg	ggcttgggtca	79680
ggaagcacag	acatcttagg	gggccaagtc	acttatctca	acaaacctag	gggtcactgg	79740
tggatccaag	gcattgcttg	cctgcctaca	ggaaaggagc	caacagatgt	gaccacaggg	79800
atgagtctag	aaccacaca	gttgttttgc	atgctcataa	aagttgaagc	aggagacagc	79860
gtgggtgga	acagtgagtg	cccttctgtc	tagaaatcag	agtttagcagc	ctgcttccct	79920
cccacgcagg	gactttaatg	tattcccaca	ctcgagtcct	gggaggggaa	aaggaagcaa	79980
agcaggagct	gacctgtttt	tatagtcaaa	cacacagaga	ggctagggaa	gcggtgaga	80040
ggagattgca	catcattcat	agctgaggtt	ggacctcagt	tgtcccttat	cccaagcatc	80100
tctccatag	acccatgtcc	cttgggtctc	tttgggtctc	cagccacctt	cacagctctg	80160
cagccagctc	accatgcctt	cctcactagg	gacccccagg	ggacatcttt	cctccctggg	80220
aatgaacagg	caggtaggag	catgagaaga	gtttccatcc	taagcctcct	gaggaaggcc	80280
tgccacagct	ccacgctcac	cagggcagaa	gccagagacc	tggctcaaag	ccacagagaa	80340
gtagacagaa	gagagaactg	ctggagccca	aggccagtga	attccccagt	tctgcaactgt	80400
ggcaccaatt	cttghtaatta	ggctgagtg	ttacaaaagag	gggtatgctc	agaagcctca	80460
gagccagggg	ttccccacca	ccactagtgg	acttaacact	agatgccttt	gttcccacag	80520
aaagagccag	agaaagtaat	ggagaaatac	tacctgtctc	atgaagtatt	gctaacaatg	80580
gtgtttaaaa	gctgatgtat	aaaaacaaca	gatctacatg	caaatttggg	ctagtctgta	80640
gctaattatg	gtgatgggtg	agcccttcta	ggctttccct	atccttataa	agtgattgctg	80700
agtcaggaat	cagcatccca	gacggatgtg	gtcattggga	agtggtgtga	agaacacgat	80760
ctaacctggc	atgtccccgt	tccaggaggc	agaggccatc	atccagaaga	tcaagcaaca	80820
gcagtattca	gaccttctgg	cccacgacca	actgtcctg	gagaggaagg	accagaaggt	80880
cttccctgcac	tttggtgagg	gcacactgtc	tctcctttgc	atcttttctc	tccatcttct	80940

ctctacctgg	acatcctgac	caaggacagg	gctgtgtctg	cagaggaggg	aagctggagg	81000
tgctagggaa	gccaaactga	tggccatgcg	tctgtttcag	aggaggaaga	gatcaccttc	81060
gccccacct	atcgatttga	aagactgacc	cgggacaagt	atgcatacac	gaagcagaaa	81120
gcaacagggg	tgagtcctcc	cagaagccac	tctcctgccc	tgteccacct	cctttaccca	81180
actcactatt	ccatgggtgg	tcctagaaag	tgggaagtat	cctacctcac	cagacaacag	81240
caaaaacaaa	agtcagaatg	cacatacagg	ggccagggat	ggaactcagt	catagagtgt	81300
ttgcttagtg	tgcacaaagc	cctgggttcc	atctcagcat	caagtccagc	atggcaccat	81360
ctatctttga	tcccaatact	caggaactag	aagtaggaag	caagaagatc	agggggttcaa	81420
ggctacccct	gactacaatg	aatttgaggc	cagtctgagc	tgcacgagac	tttgtctctc	81480
caccccccaa	aagaaaactt	tcacatgctg	acttataatc	tggtacaaaa	ctgccaccaa	81540
cttgtacatt	aaaaaataat	ccaaaaagct	gaattaacaa	tgaccctatt	ctaagatgct	81600
gagagaattt	taaagcattg	tttcttttgt	tttgttttgt	tttgttttgt	tttgttttgt	81660
tttgttttgt	tttgtttgaa	acagggctct	actatgaaaa	tcctggctgt	cctggaactt	81720
actatgtaga	ccaggtctgg	ctcaaactca	cagagatcca	tttgcctctg	ccttctgagt	81780
gctggcatta	aaggcatgca	ctactatgcc	tggttactag	taaggataat	tttatgtgta	81840
ttagtggttt	tgccatcatg	agcatctggg	gcccacagaa	gctgaaaggg	ggtgtcagat	81900
cccagaaac	agaatcnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	81960
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnntgca	82020
agacaaccag	agattttaagt	actgagccct	ctctaaaacc	cttttaaaag	acaaaaacaa	82080
agatacacac	acacacacac	acacacacac	acacacacac	acacacacac	acccacacac	82140
cacagttctg	ttagctgaaa	ctaaacaggt	tctcatgcat	ctgagtgtgg	tggtgcatac	82200
ttgtaatccc	aatacctggg	aggtctaaag	aggaagatca	taggttcaag	gccagcctgg	82260
actacatggc	tagactccaa	aacaccacca	gcaaaactga	tttcttttgt	tcaggactta	82320
gaatttcaaa	acattgtagt	tctccattga	taaagccaca	ggccaaaagta	aaccgggtcac	82380
agagtgactg	ttaatatgtt	aataaggcca	ctcatgttga	aaggaataat	gtttggggaa	82440
agatgcaata	aaatactctt	gcagaggacc	tggttctagt	tcccagcacc	aacatggctg	82500
ctcacaacca	actgtaaccc	caactccagg	ggtctaacac	tctcttctga	cctttgtggg	82560
cactgcacac	atgtgttgca	catgcatata	tgcaaacaaa	gcactcacac	acataaaaata	82620
aataaatata	ttttaaatag	tttagcacta	tctgagggtg	aattccctaa	agaaccaatg	82680
acagctgaac	ttgatgatgg	ccgatgtcct	cattataggg	tggttttgtt	tttaataata	82740
acagagaaag	cagcacatgt	gctataaaac	tgctatgtta	catttcagat	gaagtacaac	82800
ttgccgtcct	ggtgcgaccg	agtcctctgg	aagtcttacc	cgctgggtgca	tgtgggtctgt	82860
cagtccatag	gtgagtggaa	cacggtgggg	tgagggttag	gttttggggt	ctgaggacag	82920
tagcaagcca	aggggttcca	gtctgcttct	ccataagatg	agcgagtgtc	ctgaaagagt	82980
ttgtgcatct	ctatccccct	tgggctgcag	tgtaataaat	cccgtcaga	gagagacact	83040
ttaaactagaa	aggacttttt	gttgttttct	ttttcttttt	caagatttat	ttattttata	83100
tatatatata	tatatatata	tatatatata	tatatatata	tatgtatata	catatgaggg	83160
catcggatcc	catgacagat	ggttgtagc	caccatgtgg	ttgttgggaa	ttgaactcag	83220
gaccttggac	ctcaggaccc	ctaatatcat	aaacaggacc	ttctgggtact	attcttttta	83280
agagaaccag	actgagtttg	agccatataa	aagtgatatt	ttatcaggta	taaaacaagc	83340
aactaccagg	tattcacatg	gttcagccta	atgcataatc	ttaagtatgc	ccatgcagat	83400
ttggagaggc	ctaagaatat	tttattgtgg	ggctggagat	atggctcagc	agttaagagc	83460
acttgcgtgat	ctcccagagg	acctcaggtt	ggttcccagg	acactcactg	tgtgggtccac	83520
aacctcgaa	ttcagctcca	gaggatccaa	ctgctctctc	tagatcagga	gcactctac	83580
acatgcacat	gcacacgcac	acacacacac	atactttaaa	agacaaaagg	aaatcttaaa	83640
acacacacac	acacacacac	acacacacac	acacacacac	gtccaggcca	tggcactcaa	83700
cctcatccc	accacaaacc	acaggtgagc	aagtcataag	ccagagacct	ctgataggca	83760
gctcatgacc	catgctgtcc	agctgagccc	tcatgcctac	ccatccctga	gagtagagaa	83820
actggttggg	aggatttggc	cactgactag	catccacctt	atggctcccct	ggagcccatt	83880
taactccgat	cccacagatg	aaggtggttt	ctgtcccctt	catctgtatg	tgtctgctgt	83940
tgccctggga	tactttatca	catgcattac	aacccccctc	ccctgctctt	gagctctgtc	84000
ttccagaagt	ggcctcacta	gataaaacttc	atatagcacc	tgtagagggtg	tgacacagaa	84060
gcatctctta	cactcacagc	cttggccaag	gcactcttcc	gccttcagaa	cctgaaagtg	84120
ggacaggagc	tatatattgc	tcaggggttag	agaattttcc	tgccatgcat	aaggtcctgg	84180
cttccattcc	tagtacttcc	caaaccaaac	aaacaaaaac	gccagaaaga	tggtccatct	84240
cccacccctt	accaggactc	cttccctcagt	gtttcctgct	tgtctgcccc	caccaacacc	84300
tgatctcttc	agccctgtga	ccctctttct	agcactacac	agagtcagtc	atatggaaac	84360
ctaactccca	aatcctgcag	cctcgagggc	cctctcgatt	tctgactcaa	aaggtcctgc	84420
caactatctc	catcagtcag	cgagctctgc	cgagggccaa	gtgataaagg	ctcatagaat	84480
tgccactatc	tcttttagaca	ccagccttgg	ccttcagaac	tgctccataq	cacagqtttt	84540

aagaaagggg	aatgcgatgg	atcaggaagt	ggcaactaag	attgaggaga	gtaaacatca	84600
cgaacattcc	cagcatgtga	gaggcaaaaa	ggaaaaccaa	ccttatatgt	aaggctgtgc	84660
catcttccac	aggaaggctg	aagatgtata	tattttgcta	atattctttt	gagagccaga	84720
ggacgcagta	gaagggttta	aggagtgtgc	catagaggaa	gccaagaagt	tggggaaaga	84780
aaaaagaatt	catatctaga	atgcaagaaa	gggggccctg	ggacgggcgg	tgggtggcgca	84840
cgcctttaat	cccagcactt	gggaggcaga	gacaggcaaa	tttctgagtt	tgaggccagc	84900
ctggtctaca	aagttagtgc	caggacagcc	aagactatac	agagaaaccc	tgtctcggaa	84960
aaacaaaaca	aaacaaaaca	aaacaaaaca	aaacaaaaca	aaacaaaaca	caaaaacaaa	85020
aacaaaaaca	agaaaggggg	ccctgggttt	tcttccaact	ggttcataat	cgggtgtatc	85080
aactatctat	cgctgcatag	caaatgcctt	aaatcccagt	tcttaaaaca	ctgagtcctg	85140
ggctggagaa	atggctcaga	ggttaagagc	ctagattcaa	ttcccagcaa	tcacatcaca	85200
tgggtgctca	caaccatctg	taatgggatc	tgatgccttc	ttctgggtgtg	tctgaagaca	85260
actacagtgt	actcatatgc	ataaaaaata	ataaacaaat	ctaaaacaac	aacaaaccac	85320
tgagtccctt	gctcatgatc	ctttggctgc	actgaattca	gacagctcat	ttcttctctt	85380
tgggtgcagg	tggacacact	gaagtatttg	atcagatctg	gggttggtctg	actgtcagca	85440
agggtgatgg	cgatgcccag	gaggcaagtt	caggcacctt	tgcttcttgg	actcagggtc	85500
cccagactag	gtctcagctt	gtccctacag	tgagtttgte	atccacacct	tgggtcaaaa	85560
gtctcctcta	tccctactta	gaaatgcctt	ctctcccttg	caggcagtac	cagtgcatac	85620
atgacgagtg	accacagccc	tgtctttgcc	acgtttgaag	caggagtac	atctcaattc	85680
gtctccaaga	atggtaagca	atgggcaacg	tcagcttttc	ttgttttcc	caaagacaag	85740
gggcctaggg	catttgctcat	ctggttgcag	ctaccaattg	tctgggttag	atgtaggcct	85800
atcctttcct	tcccaaggcc	catggtctgc	cctacctgat	ctcttcatgt	tcaggccaca	85860
tcacaactct	agaccagaaa	accatatata	tatagtgttt	taagtaagtt	tcttctggat	85920
aagcaagtgc	ttcaagtttc	ttccaattgt	gggtcaattt	atttaattcc	atctaagttt	85980
aaaaacggac	tctaattaaa	gaaaatatta	gtccagtgtt	agtgtgggtg	catagtaggt	86040
tgagatgggtg	aaagctatgt	cgggtgggatg	tggatgaagg	aggggaggga	aactggaata	86100
aactcatgct	ttgagccaca	ggtatgggtg	caaattgccc	agcccagccc	tgggtgtctga	86160
gctacccttt	ctgcagcaat	ggctgggtgt	gcctgttaac	aagtcacatc	gggtcctatc	86220
taggtgacca	ggaagccatg	ctgtcatctc	ctctaccac	ttctgtgcag	cagataactg	86280
atttagggat	gctgggtatg	ggtaacccaa	atacagacag	agaacagctg	tctcactcta	86340
tggcctctgt	gtgaggggat	gtagacgaag	ctagacctct	gtctttccaa	ccattggatc	86400
aaacatctgt	ggatctgatg	tgacctccct	tctccacca	agtcttcaca	cctgtccagt	86460
ctccttcccta	catctggcca	cctttatagc	ctttaggctc	cacccttcc	tgtctatacc	86520
tgtcccctcc	tttgagtttt	tcagcattga	aagagttagc	tctatcacct	ccctctttgt	86580
cgggcttgggt	ttccttctgc	tgggattcag	cacacagcag	gtgcttgaaa	atatttggtca	86640
ccagtttgga	gttttgatgg	aacatttgtc	taagcagagt	ctggccaagg	tgatgtgggg	86700
tttagaagga	gagaaagaag	actaagaggg	atgggtggagg	ttgctgtcaa	gtggccttaga	86760
atttatatga	aaagatagaa	cacttgctta	aaaccagttc	aagatcagtg	tgggtgtaca	86820
cttaggtgggt	ggttctagcc	cttaggaggg	agaggcagca	ggctcagaaa	ttcagtttga	86880
gtctgactaa	actacataga	gtttgaggcc	agtttgggat	acatataaca	ccctgtccta	86940
caaaacaaaa	acaatgataa	aaaacagatg	ctgtgtaaat	aaaataaatg	tcagttaaat	87000
tagcatagtt	ataaccatga	tagctgaagg	ccagtaagtc	ccagattcca	gagtagagac	87060
tggttataac	tgtgtctgag	gaagtaagag	caggctctcc	tagctgtgga	ctcaaacag	87120
cagatttggg	tgttgagcct	ctaactctct	gccagaactt	gtggggaaaat	ttggttgtat	87180
tagtgcatgc	ctttaaccct	agcagtggag	cttaagaact	tgatgcagga	agatcccaag	87240
tttgaggcca	gccttggcta	catagtgaat	tctagacaca	cctgttctat	atagcaaaaa	87300
ggagagaggg	acagagggag	aaggagggaag	taggaaggga	aagacagaca	gacaacagag	87360
acaaagagac	agacgggtgg	ggggagggag	gaagagaatt	aattagttgg	gagactagag	87420
tgaggttgag	aggttgcccg	aggccacaca	gattattgtc	aaagctggga	ttgtagccat	87480
atcttgggtc	cttgctgtct	ttattcttcc	tgacctgcac	cctacatgaa	cagcttgtga	87540
cagaccagag	aatctgtggc	catcaacaga	gatgctgagc	ttgggtcagg	aagctttcct	87600
ttaagacaaa	tctctccttc	caacctggac	atttcccctc	tgtgggccct	cagagggagc	87660
ctggcagagt	gggtgcacag	agaggaaaac	ccagaaaaga	agctggaacc	tccatcttca	87720
gttcaggtac	agtatacaaa	taaatcctag	tctccatgc	tccaggtcct	ggcactgtag	87780
atagccaagg	gcagatcgag	tttcttgcat	gctacgccac	actgaagacc	aagtccaga	87840
ctaagttcta	cttggagttc	caactcaagct	gcttagagag	taagtgcctt	gtgaactacc	87900
ctttggggag	gttctttcac	ataactacct	ttccattgat	aactgggtga	tccctcacc	87960
agttcattcg	ttttctctct	ctctctctct	ctctctctct	ctctctctct	ctctctctct	88020
ctctctctct	ctctcataag	aatccatgga	tggacccctc	agtggtttca	gcaaagggtg	88080
ggtacctatg	gggctaccca	gtggctatgt	ggccctttcc	ccaccctgt	ggtttgttct	88140

tctactgcct	cagccctaga	gatagaggca	gccactgggtg	tccatgtccc	tgtcccctag	88200
gtccccctg	tccctgactt	gtccatcaac	attctctagg	cagaatccat	ttgagtttcc	88260
accaagctgg	gtttgagttt	ccattaagct	gggtttgagt	ttccaccaag	ctgggttctt	88320
cttgctcagt	gaaagccctg	ggcaaagctc	cagaatgctg	aagatctaac	agggcctaaa	88380
tagctgctac	tgcctgcttg	ctgttgccct	ggccctcccc	ttgccctttc	catattgtgg	88440
ctcttggtt	gtggggctca	ctgtgactgg	gagcaaggta	aagatagaca	gcccatttca	88500
ggtggcaacg	ttgagcactc	accatgctta	tgaagatcc	tttctataga	acatctccag	88560
acatactgtc	tgagactcct	atgcaagaaa	actatatgct	ttctaagtct	tcttgagatg	88620
gttcaaaag	tgatggggaa	ggataggaac	caaccttgca	aacagaggca	cagactgtct	88680
gcttgtagag	cccaggagag	ctaccagcct	tctacccctt	cacacctttt	tgcaccttcc	88740
ttcaatgcac	cccgaacctt	cctgggatta	tgccaccttc	accacaacat	cctgaagccc	88800
cagctcagca	gccaggggtc	aaacagttcc	aactggaatg	aggggctttg	tgcgggtgtt	88860
tttaataaac	tatcgagaca	tttgcataga	tttctatgga	aacagcatat	taggaggctt	88920
aaattagaat	ttcaaaaggc	tgctatcttg	agtgtcgtgg	gaagtgggac	gagctgtcat	88980
gcttcccagt	cttctcgctt	gcggtgtgat	gacttctttg	ttgcctggga	tgttgcaaag	89040
aggagtcaaa	aaggaggggg	aaatctcttt	gatttttctt	aaaaataatt	agactgtgtt	89100
ggcaaatgac	cactcttcaa	acaaaaacag	aatccccaga	aagcctgact	cctaatttgt	89160
ttggaggcct	cagagacatg	agaagggaag	gacagtcctc	cgtggtcgag	gaaggcgacc	89220
ccttgagagc	ttttctgcct	ggttctcact	ctgtggtctg	caaaccagc	ttcttcttct	89280
ctaacagggt	ttgtcaagag	tcaggaagga	gagaatgaag	agggaaagtga	aggagagctg	89340
gtggtacggg	ttggagagac	tcttcccaag	gtaatccagg	aagaaaatgt	gcctggggca	89400
gagggctgca	gcagtgcagg	ttagtacgca	gcactgcagg	ctagtgtgca	gctacagcag	89460
tgaggcgag	tattcagctg	cagcagtgc	ggtgagtatt	cagctgcagc	agtgcaggcg	89520
agtgtgcagc	tacagcagtg	caggcgagtg	tgcaagtgg	gtagtgcagg	ctagtgggca	89580
gcagtgcagg	ttggtggaca	gctgcagcag	tgcaagtgg	tgtgcagcag	tgcaggttag	89640
tgggcagctg	caacagtgc	agcgaatatg	cagctgcagc	agtgcagggt	agtgtgcagc	89700
tgcaagctctg	taggctagtg	tacagctgta	gctggtctgc	ccagagtaca	gctacagcag	89760
gttcatccag	actgtagcca	tcccagacca	aagccataga	gaccagctta	gacactgtag	89820
tcaagggact	tagggtatgc	ttttcatcat	ggatggatga	cctcaatatt	cctaagtggg	89880
tttcagccct	gctgtgccta	ctcccttttg	gttattacct	acttagtgtt	cgcagcactt	89940
gagattgaat	cgaggctcct	gcaagcacta	agcaagcacc	caaccattga	gctatgtccc	90000
tgatttccct	gaacactttt	tagcagcagg	aagttgaggg	cagtgtgcac	atccaggagt	90060
ctgtcactca	ccaagccaca	tctctagcaa	accacagaat	atttcaagcc	caactggagg	90120
tttgtgacac	atttactcag	acagggagca	ggcctatggg	gtcttctcta	atgcagccat	90180
agaccttgga	ggggtatgag	gagagggatg	gtactgcaag	tctcatttca	gacctggctg	90240
ctagaccttc	ccttggcctc	tgggatatgc	aaagaatgga	aggaagaatg	gaggcaatgg	90300
accagggacc	tggaaagctca	actcaaccca	gtttaggaaa	tgggggtggg	cagggaaagaa	90360
aaactatctc	cacttggtctg	agcaatgggg	tccacctgct	ctcttgggca	tgaagagcca	90420
agacctgagt	ttgcttaaaa	gaacatatct	agttccggtt	tactatgtat	ttattttattt	90480
tatgaatgag	tgctctatct	gcagtacat	ctgcagtcca	gaagagggta	ttaggttcca	90540
ttacagatgg	ttgtgagcca	ccatgtgggt	gctgggaatt	gaatgcagga	cctctggaag	90600
agcagccagt	gctcttaacc	actaagcctc	tctccagctc	tcctgatgca	tgcaaaccac	90660
aagactgctg	acccccatct	acctgtcctg	acctgccctc	acctgtggac	tctaactctg	90720
tgtctcctct	gcagctaaag	cccattatct	ctgaccccca	gtacttactg	gaccagcata	90780
tcttgatcag	cattaaatcc	tctgacagtg	acgagtccta	tggttaagcat	cccaggccag	90840
ggacctcacc	gtaccttagc	ctcagcctct	gaagtgaggg	tgatgggtgg	ttagcctttc	90900
aagaaagcct	gtctgccacc	acccattgtg	cctgatgct	gccacttcct	ctgtcccaga	90960
gggacttttc	ctagtgggtt	accatccctg	ctccttgctc	cttctgggtg	ccacagccag	91020
taacctcctg	cttagccttg	ctgagtgagc	tgcaaggtacc	atgtaattta	atcttagcct	91080
taagaggcag	gttctgttcc	ccagttagta	aataggtaaa	ctgaggctca	tagaggtcaa	91140
gtgactcccc	ccaggtcaca	cagatagtc	tcaactgagc	tgtggcttga	gcctgagctc	91200
gcctgtgcca	ccgtgtctct	cgtaggtccc	tagccagctc	ctcagaccca	gggtgcagaac	91260
attagcccg	ccctgctgcc	aggtagcttc	tccagctctg	ctccttttcc	tttcagactt	91320
ttccttgat	agagatttta	gactgactta	cactcttacc	tgtttatctg	attttgttta	91380
tttcttgtag	tgtttcaaac	cttgggaagt	aagctttccc	tatagcaacc	agattaacat	91440
ttaataccat	ttcttagatg	gatttttcta	ggattttaatt	tttccctcag	cattgatctg	91500
tatagagtcc	tttttccat	gaggcttcag	acatttgccc	cagattgcta	aggcatcgat	91560
tgaagggtt	tggtggcaat	ccctgcccgg	ccacttgctt	caaggagctg	tctcccagtg	91620
gtacagccat	ctctctgagc	atgctctata	tccgtctaat	gccaaagatg	tcctattgat	91680
gctctctctc	ctttccccag	gtgaaggctg	cattgccctt	cgcttgagga	ccacagaggg	91740

tcagcatcct	atctacacgc	ctctcaccce	ccatggggag	atgactggcc	acttcagggg	91800
agagattaag	ctgcagacct	cccagggcaa	gatgaggag	aagctctatg	gtaggtcagc	91860
cagcctcctc	ctggccttcc	cagaggccac	tgcactaagg	acatgttctt	tctctcagca	91920
aagctatcaa	gtatctatga	tctgtgcctt	cagaatatag	cgtggttaga	aagtctctag	91980
ggtagcgtgg	tgatcctggc	actcaggagg	ctgaggcgag	aagatcgtga	gttcaaggct	92040
agtctgggtg	acatactgag	acgatgtctt	taatgagaag	aaaacaagtc	atctcattta	92100
tctctcacia	ccgccctgtg	agtcaagtat	ggttatctct	acttggcgaga	tggagaaact	92160
ggagctagat	gactggctta	atggttgcaa	ctcaggctaa	gccaagttca	tccagggcca	92220
agcaggagct	ggtagacata	gtggttgctt	gcctgtcatg	aaaatagcaa	gtgtaggaac	92280
gcacaacatg	gggtgcaggg	gctccttatg	ccagggaggg	ctctaagggc	cactgcttct	92340
gtgcatgtga	ctctcctgtg	ccccatgaga	gcacttttct	ttctctagca	ggacagtggtg	92400
ttgaacagat	agtgggtcat	tacctagcaa	ccacattcca	gaaataaagc	ccggggttcc	92460
tttaactga	tttatagggtg	ttcccataaa	cagataagct	ctgcatggga	agtactctta	92520
tagggtaacc	taatagctgt	tctaagtttg	ttctagtctg	acaagaaatg	tctgttttat	92580
ccatggctaa	gctaacaagg	tatttcttga	aataacaact	gggtcataag	tatctgttca	92640
tctccttctg	gacccatggc	aggcatcatg	gcaggcatca	tggcaggcaa	cagttgccaa	92700
gcgtaaggat	tttgaagcct	gaaaatcttg	cttgatatac	ctctcccagt	aaccactatt	92760
tactgtccag	ctttggcctg	caaagagata	ccactttgcc	atttctaacc	catgccgtgg	92820
ctaggatctc	acatgttttc	tttcctaaat	gttctcttcc	agactttgtg	aagacagagc	92880
gggatgaatc	cagtggatg	aaatgcttga	agaacctcac	cagccatgac	cctatgaggc	92940
aatgggagcc	ttctggcagg	tagacgaagc	ttgctaagac	ttattacagc	tagctgggct	93000
gtgtgaacca	gagtcacagga	gagggtaaag	tggagttcag	gaaagccaag	gtcagaggag	93060
aagaaatggt	gtggcccagg	ggacccactc	tectacctca	gtttcacatc	ctgagttcaa	93120
atcctcatct	ctgaaagatc	aaagccttgg	agaacatttc	catgtaggaa	gggctcagct	93180
caactatcct	ttaatgaaca	cctgctatat	tcaggaagac	acagtgtaaa	tacagtatag	93240
tccatgacct	acagacttaa	caccagccag	tggaaaggatg	tgcagtgcag	gaagaaggag	93300
ctatagatga	ctgaacaatt	ccagggcaaa	agatacttct	cctgcattca	gggaggccag	93360
gaacataagt	gggtcaaaagt	caccacccat	gacttcctag	cttcatttct	gttgctgtgg	93420
caaagtccct	gattaaaagg	tagcatgggg	agggaaaggga	tcatttgact	tcccaagcc	93480
cacattacag	tctagtgttg	tggagaagcc	agggcaggaa	cctgaagtat	cacatcaaca	93540
gtcaagagca	cagagaataa	acacatgcac	ccttgcttgt	tctcagttag	ttttcttcc	93600
tcttaacacag	tgccaaaggcc	agccaaaggaa	atggtgccac	cccaatgga	ctgagtcctc	93660
ctacattaat	taacaatgaa	gacaagtccc	tatagacatg	cccacaagcc	aaacttatct	93720
agtcaattct	tcataaagac	tctcttccct	gggtgatccta	gatttgttca	agagggcatt	93780
tcaaaaccaac	caacacagag	gggtgtagcg	gggcagtcct	ttacttccag	aactcaggac	93840
gtagacgggc	tgatcttgtg	agttcacggc	cagcgttgtc	tacacagtga	attccaggac	93900
attcagagct	acataatgag	accctatctc	aagaaacaaa	ataaaacaaa	acacagaaag	93960
actaaccaac	acagtgacta	acaggaaaga	cggattggag	agtaacctgg	gcattccagt	94020
ctatgtttgtt	tcctgatata	agaaacaccc	ctgttctcctg	atgtgtctgt	gttagtcctc	94080
ctgagatgga	gcccattgtg	gaacaagggtt	tgggtgtgtga	ctacaaaagg	tgctcacctc	94140
ggctctggac	tgtctcctcc	agttttaaca	cagctgccct	ccttaccacc	tctagcctca	94200
ccctctgtgc	cgcacactca	ccctgctggt	ctgatgcctt	ccgtgggatg	tgcaaaccaa	94260
catgtcccac	gtgactctcc	cacagcccca	ggcccaccag	gttctcatte	tgctctcccc	94320
atgtgctgac	cttgtcttca	gtgaacccca	caacagggca	gcctctatcc	ctgaccacag	94380
ggccttagca	tggacctttg	cctccacaca	gagcatcctc	ctctcttttt	ctctcttctg	94440
tactccatat	aaattcaatc	cgcagacagc	cttgagtcaa	tcttctccag	ccatgttcta	94500
cctgggtgggt	ctccactctc	ctacttgtct	caacttgatc	tcttatttct	ctttgccata	94560
ttatacacgt	accagaatga	gtaaatgaat	tgttgtgccc	aagaaataaa	ggtagtgaga	94620
ggctgagaag	ctgaggaaca	ttttcctcag	gctggtgcca	gccaagaagg	gaagaggctt	94680
cagggtgtgce	tccacatctt	cctacagggt	ccctgcatgt	gggtgtctcca	gcctcaatga	94740
gatgatcaat	ccaaactaca	ttggtatggg	gccttttgga	cagcccctgc	atgggaaatc	94800
aaccctgtcc	ccagatcagc	aactcacagc	ttggagttat	gaccagctac	ccaaagactc	94860
ctccctgggg	cttgggaggg	gggagggtcc	tccaacccct	ccctcccaac	cacctctgtc	94920
gccaaagaag	ttttcatctt	ccacagccaa	cagaggtccc	tgccccaggg	tgcaagaggc	94980
aagggtgagt	tcctctgaat	tgtgtgtgtg	tgtctgtctg	agtgtctgtg	tttctgtctca	95040
aaagcatccc	ttgggggctc	ccaagggtga	cggcctgaag	agggcagagt	tgtagtagggt	95100
tctgcccact	actttggctt	ctgcctgtcc	aaaacagttg	gtgcaacttg	attttaaagg	95160
ggactgggtt	gggaacttgcc	aagaaagcca	tcttcttata	aaaactgcat	ttactcacia	95220
agacatttct	caaaccaatg	acaatctatg	ccatcctccc	cttctctggag	ttctcagtggt	95280
gaaagagggt	aggatttctg	aatagacacc	attatctacc	ccaaatctct	ctctttcctt	95340

aaaaacaagc	attcatgttc	tcatatataa	ctatctgggg	gctggagaga	tgactcagtg	95400
gttgagagca	ctgactgctc	tttcggaggt	cctgagttca	attcccagca	accacatggt	95460
ggctcacaac	catctgtagt	gagatctgat	gccctcttct	ggtgtgtctg	agaagaatga	95520
cagtgcactc	acatacataa	gtaagaaaat	aactatctgg	accctcctga	tggaaacata	95580
tattggaac	taggaacca	aatgaaggca	gagctgtcat	cctacagagg	gagccggcca	95640
gaacaggttt	aggagcaagg	accacacagc	ccagagatga	agtccaaagc	agatatggga	95700
aacattaggt	gggaatccca	ttcctacagg	atgatagatg	gccaagtga	atccagacct	95760
agttaacagt	gacacagatg	tgtctcctcc	ccagctgcct	tgagcatttt	gtggtgacac	95820
cggctctcac	attcagcctc	ttttctgagt	gctgtttctt	ttcttccctt	ttcagaactt	95880
cagttcctta	gcagagctaa	ctcatagtaa	tcagggaccc	gtgctggagc	atcaccacac	95940
gtgagggttct	cctccagacc	ctctaagtca	aatgcctata	ccaggcctgc	ctggggattc	96000
ctgtggagag	gcactgacag	tcctgtgcct	tagctgttag	ctgaactagc	tgaagggtgg	96060
gagggcaggt	cccttagcca	gactgaagtc	tacttcctag	gagcaaggga	gaaatcgctt	96120
tggcatccct	ccccggaaat	gaggatcaag	gtagccatcc	agaagggtact	gaggtacttg	96180
tttgacaaa	gcagtctctt	tcgagacccc	atgaagcaga	gctagagaca	gccacaaaga	96240
aagcacaagt	tcattggactc	agaggttccc	agagtggaa	tcactgtgtg	cttcacacgg	96300
tgaacagagc	ttgggtgaaa	catgtcctct	catgccccag	gtgacagcat	aggggtagac	96360
agctatcagg	gagcgagcta	caggactagg	tggcaatagc	cacccatccc	aggttcccca	96420
aggctgcccc	aacttttagca	tttaaaagtc	ccatctcctg	gaaaacactg	cagtcctggt	96480
aagcttgagc	caccaccatg	tgaggtcaag	gctgtgtgcc	aagaaaggag	actctgctca	96540
ggccctagcc	ggcatccatg	gtctcctgca	caagaactga	ctctgcccc	tggataagag	96600
cttggtttcc	tgttgccat	actcaaactt	tttttttttt	ttgactaaaa	ctcgtgagtt	96660
taatctgttt	ttcctaacct	ccattaggt	agcaggtgta	ccctgaacac	tgttggaagg	96720
atcactttct	agcttctatc	tcattggctc	catcctctgg	cctctctatc	tctatagctg	96780
agtggatagc	cagatgccac	acacatgcac	aggcacgcac	acgcacgcac	atgggggggt	96840
ggggagatgg	ggggggagtt	gcttgcttct	gtctaccaca	atgccctcag	aggccagaag	96900
aggacattgg	attcccaaga	actgcagtta	catacagttg	tgagttacta	tgtagatgcc	96960
aggagtcgaa	cctggtaact	ctggaagaag	agcaataact	cttaatcaca	gaaccaactc	97020
ttcagcccgg	catggtgttg	catgccttta	atcccagcac	tcagcagaca	gtgaggcggg	97080
cggatctgtc	agttctatat	caaccaggga	tacgtagtga	gccctggtgt	ttaaaaagaa	97140
aaattctcaa	ggcagaatcc	atgtaaaaat	tattccctgg	aaaataagta	attcggaggc	97200
atthctgtgg	tgcattgtat	atgatcacta	ctgtagttag	ttccagagca	ttcttatcac	97260
ccctaaatga	aaccacagag	catgaagcaa	ctactcccca	tgtcctgact	tctgaacata	97320
tttctccccc	atctcctctc	cttttcccca	tggttccaga	cctggggatc	tgggaaaggt	97380
ggaagctctg	ctccaggagg	acctgctgct	gacgaagccc	gagatgtttg	agaaccactc	97440
gtatggatcc	gtgagttcct	tccttaagct	ggtgcccagg	aaagagcagg	agtctcccaa	97500
gatgtgcgg	aaggagcccc	cgccctgtcc	agaccagga	atctcatcac	ccagcatcgt	97560
gctcccca	gccaagagg	tgagagtg	caaggggaca	agcaaacagg	cccctgtgcc	97620
tgtccttggc	cccacacccc	ggatccgctc	ctttacctgt	tcttcttctg	ctgagggcag	97680
aatgaccagt	ggggacaaga	gccaaggga	gccaaggcc	tcagccagtt	cccaagcccc	97740
agtgccagtc	aagaggcctg	tcaagccttc	caggtcagaa	atgagccagc	agacaacacc	97800
catcccagct	ccacggccac	ccctgccagt	caagagtcc	gctgtcctgc	agctgcaaca	97860
ttccaaaggc	agagactacc	gtgacaacac	agaactcccc	caccatggca	agcaccgcca	97920
agaggagggg	ctgcttggca	ggactgccat	gcaggtatgt	tgagctgtat	gtatatgggt	97980
gtatatgcat	atgtgtgcac	gcatgcatct	gtgtgtgtgt	gtatgtgtat	gtgtatcctg	98040
ggatgtgctc	tgtgaagaag	ggctaaacct	gggttttgta	cctggcatct	gttcctgttc	98100
ctgactggtc	ctacagggca	tctgcagtga	ttcttgagag	cgacagagag	gagctgttga	98160
cacctgtaga	gtaaggaggt	ctaagtcatt	tatatthttag	tcttatcatc	cttaattagg	98220
gtatgcactt	aaaacattcc	tgggtgtttta	gagacctcag	gcagaatatc	cccttcctat	98280
agaatagatt	tattgtagag	caaaactaac	aatgacttaa	atacacacac	acacagagtt	98340
gagactggac	atgatagtac	accgctataa	tcccagcaca	ctgtacatct	agagcaagag	98400
gattggaagt	tcaaggcaag	cctgggctaa	atggtgagac	agacaacagc	agcagtgaga	98460
gagctaaagt	tatatagctc	agtgggagag	tgcaaggaa	gcctccagaa	agcccttagg	98520
tgaatctcca	gtatcacaag	gaacagagtc	tagaactaga	ctgctggcca	ttccattagc	98580
atacttattg	cagcattgag	ggtcagactt	ggggccttac	actaggcaag	cactgtacca	98640
cagtccctgtg	ttggatttaa	atcccaagtc	tccactgcac	ggctctatgc	cttttttaaac	98700
tgtcaaaaag	aaactactag	tgccagccat	tgtgtagatt	tatggggaga	aaaagattat	98760
gccatatact	gttattgttt	gtgatcattt	ttacagttat	cactcatttt	gaaactatat	98820
acaacatata	tctatctcat	aggtactgtg	tttgtaatat	atgcatgaat	taacatataa	98880
ggcacttagg	acagtggtggc	tctttcacat	taattagcct	gacctaactt	acagcactag	98940

```

gggtcaaatac tggcaaggcc ctgatagccc ttctctgttct catcttacag tgagctgctg 99000
gtgatcggag cctggaggaa cagcacaaag cagacctgcg cctctctcag gatgcctctc 99060
tcaggatgcc tcttgaggaa cctcctgcta gctcttcttg cctagcttca agtcccaggc 99120
tgtgtatttt ttttcaggaa acggcctcac ttctctgttg tccaagaagt gtgctgctgg 99180
ctgccacact gtgcggcaga tgctaaagct ggatgacaaa cgcacgccat acagacagca 99240
gacagcggca ctgggtctca gaacttggat tcctgggcct tcttcagtc gccgttttaa 99300
agaaaaggaa taacggagct gctcatccga ggggtgaagat ataaataata atattattaa 99360
taataataac agtcagggtc catgtgctgt gtttaagtgt ttatgaacat ttgtcgggct 99420
ggcctccagt gctgaggtgc cagtcagcct gaacctatg cccaggccca ctaatcccaa 99480
atgggtgggtc ctgagatgtt tttaaaaagc attaaagaaa accatcggtc tcttagagct 99540
aaccggccgg gctctactgc agggaccoga acagtctgca tggctaagtgc gcacaaggag 99600
cctggccctg tccagcttca gagatccaag ctgctttttg ctgggggttct gtacaggcc 99660
tgatcctctt ggtttttatg gggtttcaag tctgccagag tcagaaatca gctctaactc 99720
gccagtgaag agatctggcc ttaacttaag ccagccacgt caggcccctg ctgagcctat 99780
ggaccaataa atactccccg tgccactgga ggtgggcagc tatcaccata ccctgagttg 99840
ggccaagccc accccacccc tacctcgcaa catctctgat gtactgagga agagtctcca 99900
ccatagtccc caagggtgca gttctccagc ctgctatcag ggaagggtgag cattgggtccc 99960
aggctctcaa aatagtgcag cctcttcttc ccaagctctg ggggtgcacc tgtgtccttg 100020
gttaccagga gactagggtt gtgatattct ttctgtctt gctttttgat ataccaggat 100080
taatgtagga aaccagacct agattattca ggagagtagg tatatccctt gtgtttccca 100140

```

&lt;210&gt; 2

&lt;211&gt; 11928

&lt;212&gt; DNA,

&lt;213&gt; Mus musculus

&lt;400&gt; 2

```

gcttcccag gagcacctga aagccatcca ggattatctg agcactcagc tcctcctgga 60
ttccgacttt ttgaagacgg gctccagcaa cctccctcac ctgaagaagc tgatgtcact 120
gctctgcaag gagctccatg ggtaacggag agccctgaga gaggggtggg ggagcttcat 180
gggtaatggg agccctacc caccaggag gatggccaca gcaagagaaa gtgctcatta 240
gagtgaacct gggctctctc tctgtccaga tgtctctgca gcactcacag taattggccc 300
agggtgagtc tggaaatgtc caggcttgtt ggaagctctt gctctcatag aatctgagct 360
ctaactgagc tgggaaagt gatcatttgt ttattccttt tagggatttg ggggggcacg 420
gatgtatgca tgctggtatg tatgcatgct ggggatgtag ggcctcatgt gtgctaaata 480
catgtgccta ttctgtaatg tttcttgtt tgtttttaac tcaaattaat tagaggcagt 540
ttctctgttt aaaaaaaaaa ggtggaagac actgccatct catttgtgtt gggacttgag 600
atatatatat atatatatat atatatatat atatatatat aatttattta tttatttatt 660
tgtgtatgtg tgtataagaa agcgggggac actcatgtgc catggtgtgt atggggacac 720
acgtgtcctg tgggggaaca ttgtgtcagg ggaacaagca tgtgctgagg gtggggggat 780
atatatgtgc cactgtggtg atttggttg gggagacatg tatgtgctat gatacctgtg 840
tgtaggacag aggacagact cacacaggag ccttcttacc tgctgagctg tctcagcagt 900
ccaagccctc caggctgtag ccactcttcc tccttgctac agtcccacat ccagccaaca 960
cataaaggct ctggcaggaa ataaattaaa ttgtcttgt gtgtgggtgc tgacggctca 1020
agtcttccgg tgatggtaat gtttaaagca agccaacac agttctccag tgccgaagt 1080
attgaatgac tgaccaatgg gtaacactta ggatttttaa aaattatttt gattgtacat 1140
tttgaattac attaatctta aaataaacta caaacataaa cacactgtgg ttggagctgg 1200
gcatagtggc acatgccttt aattccagaa ctggggaggc agaggctggc atatctctat 1260
gactttgagg ccagcctggt ctatataatg agttccagga cagagagatc ctgtctcaca 1320
aacagagaaa accccataac tatactttta ttgtactggt tggctctacag tgtaaagaat 1380
tggcaaagaa tatgaaagat tacactggga agaaactgaa agccatccag agtaatgaag 1440
caaaccctc acaaatgtgg gaaacatttc acatgggtga gggcttgca cagcctgtg 1500
ctaatttatg ttttggtacg tgagcactta gaggagtttc cagttctctg gtgctctacc 1560
aatcttagct tgggttaatat tcagggatga gtcttccatc accggtaatc taatttgcca 1620
ttctatttaa aggtctctaa aggcacaggc agtgcatagg gtaaatgtgg caaataattt 1680
ctttgacaaa tcgaccaatt gtcagattgg cctgctagtc atttgtttca atgagaactg 1740
ttttttctca aaggatgctc ttgtacaccg gctagaagca gggctgtcat ttttatagg 1800
ctctgtggta ttttgttgtt gttgttctgt tgttgcaaat catttatcac tgagggaaaa 1860
tacacacaaa gggcccttcc tttaaaagta tacatgtatc attttgtgac agctcataag 1920

```



```

aagctgtttt tttctgctg gacacaggtc ctgacctgtg ctgtgtcctt gctaagcttt 1980
gtcagaccct tccacagctc cccccaacaa cgagttcccc agtacctgcc tcacctcatc 2040
actatggtga cagcagcctc tgatgcgcct gactctctgg cacattatgg cagtgttaaa 2100
agcttccatc tctctctttg ctgaataatg aacctcaggt tgttcaagag accggaatgt 2160
tcttcacctg cctgcacaca tctcttcaact ttctttttata gatcaggtag ggactgggag 2220
tgtagatgga acaaactggt ttccgttccc cagccatctc tgcaggtgca ctccacataa 2280
atcaagtgtt aaaagtgtt tgattaaaca ggacaggcgc gttcttgagt tcatctgttc 2340
acatactgtc tggcaagcgc tgactgaggg tctcctctgt accctgttct gagaactaac 2400
aaaagacgaa tcaacataca gaaaactggt atttagtga tgaattaaact aacgaaggca 2460
tgggctggag aaataactca gcagttaaga acatttgctg atcttgaga ggacctgggt 2520
tgggttctta gcacacacag ggacagttcc agtcccggtg tgtccttttc tcacttctgt 2580
ggacacaagt ttacacata gtgcacacac atacactcac atatataaaa cagaacattt 2640
aaaagtatgt ttaaataacg gaatcattta tataggtttt catttacata ggtaaatagg 2700
caaaaatctg catttttattg ttcttaagtt ttaatttatt ttctctgtg tgtacatacg 2760
catgcctcct tatctgtatg tgcgtgcact gtgtgcatgc atgaaccac agagaccaga 2820
agagtaccac agattctctg gagctggagt gattgatagg ctgttgggag ccactccaca 2880
tggggattca gagttgaact tegtctctg caagaacagc cagctcttaa ctgatggctt 2940
ttacctccag ccaccttttc ctcatTTTTT aaatttctt ccttctttt tgagacaggg 3000
tctcaatact tagctcatcc caacttgacc ccactcttct cttgccttag tcaccacaat 3060
gttttagttta taagcatgcg tcactatgcc cggttttaaa taaactcacc cataatccca 3120
gcaactgaagt agacaaaagg gaggatcgat ggggctgact ggccacaagc cgtgcttcaa 3180
gttcaatgaa gacctgtct caagggaata aggcacagag gatagagcca tacgctgac 3240
ctctctctct ggctcttacc caggcacatg tgcatacaca caccacacac acacacacac 3300
acacacacac agagagagag agagagagag agagagagag agagagagag 3360
agagaaactt ttctctctt ttttttaaaa atattttta ttcatgtat atgagtacac 3420
tgttgctgtc ttcagacacc ccaaaagagg gcatcagatc ccattacaaa tggttgtgag 3480
ccaccatgtg gttgctggga attgaactca ggacctctgg aaaagcagtc agtgccttta 3540
accatctctt cagccccaca aagaaacttt taatgagcaa ataattgctt ccaagtaaat 3600
actactaata tttttctaac catactatac aaggaattat taaagaacgg ataattaggag 3660
aataaaaaat tataagtac tttataatgc tatctaatac atctagaaca aaaacactgt 3720
aataatgcaa aagagcgag tgcctagatt aaataaataa aatgcagacc aataagtaaa 3780
ctttatagca gcacatggaa atgacgaaat tcctaacaaa aagctcaaga tgggcagttt 3840
atttaaagtg aaatacagga gaaataaagc acagaaagat actcaaaggc atagaagta 3900
acataggggg gctggcgaga tggctcaaca ggtaagagca cccgactgct cttccaaagg 3960
tcttgagttc aaatcccagc aaccacatgg tggctcacia ccatccgtaa tgagatttga 4020
ctccctcttc tgggtgtgtc gaagacagct acagtgtact taacatataa taaataaata 4080
aaacttttaa aaaaaatta aagaagttaa catagaagcc cactcaggac ccactcagt 4140
cctagagtat gacattatta tggacattaa aaagagaaaa ttcagcagta gtgtgcatgc 4200
actgcatata tacacaaatc cttgagtttc ataccaaatg ccttttagacc acttgtggct 4260
ctgcaaacct gtaatcctag cacttgtgaa aaggtcagct caggaacttt gggaaggtca 4320
tgaaactctt gccctccag aaggagagg ctaattaaca tttctcagac cacagggcgg 4380
gaaccgacct gcgggtgggg acagactgtt gccatttcc agactaggga agtccttgtc 4440
acctcattcc ctaaagacca atcaatttaa aggtgcact gttccgcaa tcatattgtg 4500
cctagtgtg gatgtctat tctgccctta gaaaccgtat aaaaactagc gaaggggtac 4560
caggggtaac cccctctct tcaagtctgg gacaatccca ctacactgga acaataaatt 4620
cctcttgctt tttgcattga tcacagctcc acttcgtggt aagctaagac tccctggagt 4680
cttacattgg caaatgcagg caaaagaatc cgaaactcaa ggtcatctaa aactacatag 4740
caagcatgct gctagcctgg gctccatgag accctggggg aggggcagag ggagaccgtt 4800
cagaagacag tcaagatggt gcagcagcac aggcagcctg gccaccagtg ctgtcaccag 4860
acatgttaat gttggaataa agcctcaatc atgactctcc cagttttata attggaaata 4920
agaaaggaaa gactatagga acaactgtgt tcagaacact atttataata gcaaagatct 4980
cagagtaacc caaacttcta gacattgatt tgggaagatc tcttggcagc ttattttgaa 5040
aactttacaa tgttaaatat gtaaaaacaa ggacagtttt gttttttgtt ttgttttgtt 5100
ttgttttagg gatatatatt catatatgta tatgaatgaa aaccctaaact taaaattccc 5160
cactatgctt taaaggcttt ctgacaataa cagaaagaga aatagagaat ccataaaaac 5220
tagttctgaa actatcaata ggcttgacac tcttttagctg ccaggagagc tgaatctgaa 5280
cacagggaac cccaccagc accccaaatt tggattattg ttttatttta tctttccctt 5340
accccaaga cagggtttct ctgtgtggtc ctggctgtcc tggaactcgg agatcctctg 5400
cctctctgcc tctctgctc tctctctctc tgcctctctc tgcctctctc tctctctctc 5460
tctgctctc tctctctctc tgcctctctc tgcctctctc tgcctctctc tgcctctctc 5520

```



```

tgccccctctc tgccccctctc tcttccccctc tctgcctctc tctctgcccc tctctgcctc 5580
tctgcctctc cctcctgagt gctgggattt aaaggcatca gccatcactt ccagcttctc 5640
ttatcatttt aaaaagaatt tcctatgtga ctactgtatt taaatcacca caggccaat 5700
actccccccc aactcctccc aaatccccctc taccactca aattcttctc ttgtattctt 5760
tatcattatt atacatatgt gtatatatgt gtgtgtgtat atatatatat actataact 5820
gctaattgagt aacatttagt gttattcatt gttgcatgtt ttcaatgtgc tttccaggag 5880
gctgggggga tggctcagtg ggcaaaattc tagctgcaca agcctaagga ccagggttca 5940
gatccccaat ataaaggctg gctggacatg gtggcttgcc tatgatacta gcatgcttgc 6000
tggaagcaaa gacgggaat ccctggagac ttagaatctc agaagtgatc tgggctggac 6060
agactagctg aactggccag ctctgggttc atcaagaaac cctacctcca taacataaag 6120
tgtgatggag aaaggcacct aatgtcaacc tcaaaacctt acctgcatgt gcacacacat 6180
acatccacac cacacacaca cacacacaca cacacacaca ccacacacac acacacacaa 6240
ataaataagt aaataaataa aatatttagc tctccagacc aaatcttggg gaaacctatg 6300
catttgcatt tgttgtgtgc ctacaaacac tgaagggtta gaagcatgct ccttagtaat 6360
tttatagcag tttgcgtttc cagattgaaa acagattcta taggtacac agtgctaaat 6420
ggattatgct cgatacaga ttgaaaagga tacagattga aaagggtcgg ggtctgggccc 6480
aggatgacgg gccaaactgat ctttgcggg gcttgtcctt cagggaaggg ttacaggatt 6540
caccactggg gtgtggccta tctgtgttta ggactgaaat tgctggagt gtttctagt 6600
cccactagtt gttgaacttt acctgaacc tctgtctcca ggaagtcat caggactctg 6660
ccatccctgg agtctctgca gaggttgttt gaccaacagc tctccccagg ccttcgcccc 6720
cgacctcagg taaggggttt ggatttggaa agatgcaatt gctataggag ggactctgaa 6780
ggcagacaga cgcaccgcct cctcacgttg gctagtctaa tataaacatc gcggtggatg 6840
gtgaggatag actccatgcc cttttgtgaa ggcatttctt ggcatcagct cctgacttca 6900
gacagtttca cccatcagac aaattgcctg gtgttggagg aggaggtgag cagggccatt 6960
cccatcattt ctctcagaa atggaaaaggc aaggaaaaca tgaggttctt cagacactta 7020
atccctggga ctgcaaaatg gtggtgcccc tctccacag ctgctcagg cggggcagga 7080
gatgagggcc aaatgaagca tagatctagc tatttttttt ttagtgctt cagtaaattt 7140
aaaatcaaat aaggggaagg acctagatct ttatgttatg gcattgttaa aagtgagaac 7200
ttgtagccag ggtgtggtgg cgcacacctt taatcccagc acttgggagg cagaggcagg 7260
cggatttctg agtttgaggc cagcctggct tacagagtga gttccaggac agccagggct 7320
acacaggttt ctctatagc agaaacctgt ctgaaaaaac caagaaaaaa aagaaagaaa 7380
aagaaagaaa gaaagaaaga aagaaagaaa cgaaggaagg aaggaaggaa ggaagaaaga 7440
ggacaacatg gtctaggggt cagagagcag aatctccaaa aacaccaaca atgctgctg 7500
taaattgatg tcgttgattt ggggatgttg gctccagct caccatttcc tgctttagcc 7560
tccaaagtgc taggattata ggcttgagcc aacacatctg gcttacgcct attgtgtgtg 7620
gaaggggagt gctgagtgtg ctctgtgtt tggacttat atatgaatat atgtatatac 7680
gcatgtacgc atactgcat gtgaaggcca gaggccaatg tcagctgctc tcatcttctc 7740
ctttttatta cattgtattt atttgtttgt ttgtttgtt gtttcatcgt acgcatgcag 7800
ccactcatga gcatgtaaca gcacaggtat gaaggtagac ttgcaggagt cagtctctc 7860
cttctgtcac ttgagttcca ggaccacact ccagcccca ggcttgggct gtaagagagc 7920
catcttactg gtcctctact ttgtctctg agatagcatc tagactcacg gaacctggag 7980
ctcatctaga ttacattgg ctggccagct gatgcatttt aaggtcaaat cttcattcca 8040
tccctacccc acttccactc ccagtgtctg agttcgggac acctgccacc aagcccagtt 8100
tttctggat gcagaagctc caaactcagg ttcccatgtt cgcattggcag gcacattttc 8160
agttaagcct tccccccagc tcctttaccc tggctctctga atggggggga ggctataaat 8220
caggctgctc tcagacatta ggtaggaaat agaccatata catgaggaaa gatattcacc 8280
tgccccatgg gtaccaggaa gtgatgtcca actcctctt tgccttctcag gagaaatgct 8340
gactactacc tctggttaatt ttgatgttg gaggaacagg gacattcata ggacccatt 8400
cctcgctggt gagagtggag acaggttttc tgaagggcag gagatctgtg tagaaaagat 8460
ggatgctgtt tctgaagggt aaatggagggt agagtgcacc tgggagagag gggaggtggg 8520
ggatggttgg gaggaatgaa aggaagagag acggcagttg ggatgtattg tataagagaa 8580
gaatcagaaa gaaaaaaaaga aaagctacct gcacccttca agtgttctc tgtgtgggag 8640
gctgtctcag ggaactacat ggacccgaga ggcactcagt agggtaggta cttgatgttg 8700
tgtccctgaa aacaaggaca ggaaatctgc tgcattggcct aagatggcaa aatgtggc 8760
aatcaagtaa ggcccaggat tctgtctgtg gtgcagacct gctgtagaat gagctcccag 8820
cattcccact tgtgtgttg agacagcatg ttgcagagcc atgtgaggat gagggccag 8880
gccaggagga tgtcaaccca ccaccatgta gccagtgggc tgggggagct tggggccacc 8940
aaggagcttg agcagactga cagtgggtta tgtacacaag tgggcgtgtc acacaaccgt 9000
gcaacacaga gaaaatccct gtgatgacaa cttctaaacc acctgaggc aaaaggagta 9060
gacaggggat tagagcctag catattggag tgcagtggcc atgcagctct tggaaagcgt 9120

```

```

aggaaggaaa tttcctggaa ggatagggtg tcttcctage agcctcgtca atagatgtca 9180
atgtatgagg tagtacctgc tacaatcctg cttcttcaga agactgaggc aggggggatta 9240
cttgaaccca gaagttctag gccagtatgg acaacatagc aagagtctga tttaaaaaaa 9300
aaaaaaagta aagagggaaa ccaaataggt gacgtgccac actagtgtct tcctgtctcca 9360
aggggtcctgg gcacatgagc ttgcttagtg ccagaaaagt cagaggagga gagggcagac 9420
agagaccctc gtctccacct cctttgactg actaatgggg ctggatataa tctgttttac 9480
aaaaggacag cttttcagag ctgtttctat ctaagggttg tctgaatagc catctcgaaa 9540
tatgccagag aagaatattt agaggcggcc tattttgggtc tcccacaaag atttcacagg 9600
gaaaatgtat ttgtgttcta tttatacaac taaaaatatg catcagcccg gggaactgg 9660
ctctttgtcg cctttgaagt gaaggatggg ttaatttcta agaaagtaaa agcaaatgta 9720
gtgcaggcac cacggatgct gccagacacc agcgtttaag tggcttgaat ggaagagcac 9780
acccaagta tttgaagtag gtgaggagag agaccagggtc tccagagttg ggctgcagt 9840
ggccagggtg agtccaaggg gcagtatgac gcagaacaga gtggcaacct ctaccagtag 9900
tagaattcag gtctcattcc taacctccca taaagcagag aatattgcac tctctttctc 9960
tctctctctc tctctctctc tctctctctc tctctctctc tctctctaac tcacagagat 10020
cctcctgcct ctgcttctctc attgctggaa ataaaagcat gtgccaccac atgcagctct 10080
ccttgtttg agagagaaa gacagacagc agacagacag acagacagac agagtataat 10140
gtgtgcagat gtccacagaa ctccgaagag ggtgttgag cccagaaaac cggagtctta 10200
ggcagttgta agctgtccag tgggtgttgc gaactgaact caagtcctct ggaaaactgg 10260
aaagtactct taaccactaa gctatctgtt agtcccccaa gaatgtctta tcttgatagg 10320
ccttcagatc tcacagtcca gggggctgac ctgctacaag gggtagagga aagaaaactg 10380
gctccaggcc tgagttcaga gtgacatctc cagagttctt ccttcccttg ccgtgtagta 10440
atctcctacc ctacacgagg agaaaaggaa cagatatgtc cagtatgcct ggcattctga 10500
aagggcactg actcttgccg ctgtagagcc cctttccttg gggagtgcag agaagtgtg 10560
ctagagaggt tcaaaaagaaa cacaacagtc aaaatagttg ctggccaagg gagggcatgt 10620
gccctgtatc gggcctgcaa agcccttca acagtgtagc caaggccatg tttgacagta 10680
caagcctgta agcccatgca taagcaaggg ctgggataaa gggctactgt tcaagttagt 10740
tatatacaca tcaagtttgt tcatcttaca tccttaggta aggggtgggta gcttgtagc 10800
tccacctctc cagcaggaaa gcatgtccac ggtaaggtag atactgtaga gtttagcttt 10860
gttttgacct ttctttcttc gttgcttggg attgtacttg ggaccttcac tcatgcagcc 10920
catgtttcag ccttgcataa atctataaat atatggttg agtcagggtcc agcatatgtg 10980
ggtaccagtc ccagtctggg atagcaattt attagctatt tgtctttaga caagtaacta 11040
tagggttctg agctgtaatg gccaccaaac aggtccttta agtatcttta ttggccaaac 11100
accagggtta aaacactgaa cacagtatgg tggtttcaac tgtcatggag tccacattct 11160
agcaggggaa ttttacacac aactcaciaa tgatacagac cacagcaggg tgatttgaca 11220
gagacaacga agggcttgcc tgagctgcag tagtccggga gcagctcttg atgggatctg 11280
agatctagat ggcagagcgc ccttcctaga tgaatccaa gcctctggtt actcctcagc 11340
aagagttagc ggaggaagac cagaagttag gaggaagggt ctcagagtgg agcatctaag 11400
cctgcagcac ctccagagac aggccttggg gcttgctgt ggttcagcct actgtgggaa 11460
gagaccagtg agggttttta gctgtgggga tgggtatatt ttggaaggcg aacaagaaca 11520
ccagcagctc aggcattgta ggaggcggg aagggtgatg atgatcgagg acggccccag 11580
agatttttaag gactgtgtag gtggggtgag aagcactgtg ggtgaggtgg agttagaggg 11640
aagaagtga gatctatttt tttggccatg aggaatttgg ggtgcacaca cacacatata 11700
tacacacccc atcatgtcct tgtctacaga cgggtgata atggtactgg aaggtagaag 11760
gtggaaaagag gaccagcaag gaaggatcc agtggccacc ccacagctca cctcagaca 11820
gaccttggtc tcattttcat taccagggtg cccggagagg ccagtcccat caccatgggt 11880
gccaaactca gccaatgac aagtctgctg tcttcattg aagataag 11928

```

&lt;210&gt; 3

&lt;211&gt; 560

&lt;212&gt; DNA

&lt;213&gt; Mus musculus

&lt;400&gt; 3

```

ctctgggttc atcaagaaac cctacctcca taacataaag tgtgatggag aaaggcacct 60
aatgtcaacc tcaaacccct acctgcatgt gcacacacat acatccacac cacacacaca 120
cacacacaca cacacacaca ccacacacac acacacacaa ataaataagt aaataaataa 180
aatatttagc tctccagacc aaatcttggg gaaacccatg catttgcatt tgtgtgtgtc 240
ctacaaacac tgaagggtta gaagcatgct ccttagtaat tttatagcag tttgcgtttc 300

```

```

cagattgaaa acagattcta taggctacac agtgctaaat ggattatgct cagatacaga 360
ttgaaaagga tacagattga aaagggtcgg ggtctgggcc aggatgacgg gccaaactatc 420
tttggccggg cttgtccttc aggggaagggg tacaggattc accactgggg tgtggcctat 480
ctgctgttag gacctgaatt gcctggagtg tttctagttc ccactagttg ttgaacttta 540
ccttgaacct ctgctcccag                                     560

```

&lt;210&gt; 4

&lt;211&gt; 560

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 4

```

agcccgagga ttggagttgt caatggatgt gacaatggga aaatccgtct gagcctgcat 60
ttgggctgct aggaggggat ttgcatcaga atccacagat caccagcact gggcagccct 120
aatattttaa atgcagattc tagactcaat caggcgaggag ccagaaatt tgcattgtta 180
acacctgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgttttata aacacagaag 240
gttgggaacc atggataact aagtgaagtc attttgtcac tcagatttga attttctaca 300
ggctatagag tgcagtttgg ctaaagcaaa acctaggtag agtcaggact acacaatttc 360
agttcgctgt ggggtgggaa gggatgggtg ggccagtgtg ggcaagcctt gatccttgcc 420
cgggctgtgt cttctgggga gaattacctg cttctgtgtg actgaggggt ccctcatctc 480
tggctagagc ccgtgctgcc atggaagact ctttccggtg cccactaatc cttgatgttc 540
accttgtccc ctgccccag                                     560

```

&lt;210&gt; 5

&lt;211&gt; 11539

&lt;212&gt; DNA

&lt;213&gt; Mus musculus

&lt;220&gt;

&lt;221&gt; modified\_base

&lt;222&gt; (1847)

&lt;223&gt; N = A, C, G OR T/U

&lt;400&gt; 5

```

gagctctaac tgagctggga aagttgatca tttgtttatt ccttttaggg tattgggggg 60
gcacggatgt atgcatgctg gtatgtatgc atgctgggga ttagggcct catgtgtgct 120
aaatacatgt gctattctg taatgttttc ttgtttgttt ttaactcaaa ttaattagag 180
gcagtttctc tgttaaaaaa aaaaagggtg aagacactgc catctcattt gtgttgggac 240
ttgagatata tatatatata tatatatata tatatatata tatataattt atttatttat 300
ttatttgtgt atgtgtgtat aagaaagcgg gggacactca tgtgccatgg tgtgtatggg 360
gacacacgtg tcctgtgggg gaacattgtg tcaggggaac aaacatgtgc tgagggtggg 420
gggatatata tgtgccactg tggtgatttg gttgggggag acatgtatgt gctatgatac 480
ctgtgtgtag gacagaggac acactcacac aggaagcctt ttacctgctg agctgtctca 540
agcagtccaa gmcctccatg gctgtagcca ctcttcctcc ttgctacagt cccacatcca 600
sccaacacat aaaggctctg gcaggaaata aattaaatyt gctttgtgtg tgggtgctga 660
cggtcgaart cttccggtga tggtaatggt taaarcaarc caamatcagt tctccagtgc 720
cgaagtattt gaatgactga ccaatgggta acacttagga tttttaaaaa ttattttgat 780
tgtacatttt gaattacatt aatcttaaaa taaactacaa acataaacac actgtgggtg 840
gagctgggca tagtggcaca tgcctttaat tccagaactt gggaggcaga ggctggcata 900
tctctatgac tttgaggcca gcctggtcta tataatgagt tccaggacag agagatcctg 960
tctcacaacc agagaaaacc ccataactat acttttattg tactgggttg tctacagtgt 1020
aaagaattgg caaagaatat gaaagattac actgggaaga aactgaaagc catccagagt 1080
aatgaagcaa acccctcaca aatgtgggaa acatttcaca tgggtgaggg cttgcagcca 1140
gcctgtgcta atttatgttt tggtagctga gcacttagag cagtttccag ttctctggtg 1200
ctctaccaat cttagcttgg ttaatatcca gggatgagtc ttccatcacc ggtaaatctaa 1260
tttgccattc tattttaaagg ctcttaaagg cacaggcagt gcattgggta aatgtggcaa 1320
ataatttctt tgacaaatcg accaattgtc agattggcct gctagtcatt tgtttcaatg 1380
agaactgttt tttctcaaag gatgctcttg tacaccggct agaagcaggg ctgtcatttt 1440

```

tatagggtctc tgtggtatgtt tgttgttgtt gttctgttgt tgcaaatcat ttatcactga 1500  
 gggaaaatac acacaaaggg ccctttcttt aaaagtatac atgtatcatt ttgtgacagc 1560  
 tcataagaag ctgttttttt ctgcctggac acaggtcctg acctgtgctg tgtccttgct 1620  
 aagctttgtc agacccttcc acagctcccc ccaacaacga gttccccagt acctgcctca 1680  
 cctcatcact atggtgacag cagcctctga tgcgctgact ctctggcaca ttatggcagt 1740  
 gttaaaagct tccatctctc tctttgctga ataatagaacc tcaggttgtt caagagaccg 1800  
 gaatgttctt cacctgcctg cacacatctc ttcactttct tttatanatc aggtagggac 1860  
 tgggcgtgta gatggaacaa actgttttcc gttccccagc catctctgca ggtgacactcc 1920  
 acataaatca agtgttaaaa gtgctttgat taaacaggac aggcgcgttc ttgagttcat 1980  
 ctgttcacat cctgtctggc aagcgtgac tgagggtctc ctctgtaccc tgttctgaga 2040  
 actaacaaaa gacgaatcaa catacagaaa actgttattt agtgactgat taaactaacg 2100  
 aaggcatggg ctggagaaat aactcagcag ttaagaacat ttgctgatct tgcagaggac 2160  
 ctgggttggg ttcctagcac acacagggac agttccagtc ccggtgtgtc cttttctcac 2220  
 ttctgtggac acaagtttta cacatagtgc acacacatac actcacatat ataaaaacaga 2280  
 acatttaaaa gtatgtttta ataacggaat catttatata ggttttccatt tacataggta 2340  
 aataggaaca aatctgcatt ttattgtttc taagttttta tttatttttc tctgtgtgta 2400  
 catacgcatg cctccttctc tgtatgtgct tgcactgtgt gcatgcatga acccacagag 2460  
 accagaagag taccacagat tctctggagc tgaggtgatt gataggctgt tgggagccac 2520  
 tccacatggg gattcagagt tgaacttctg tctctgcaag aacagccagc tcttaactga 2580  
 tggcttttac ctccagccac cttttctctc tttttaaaat ttcttctctt ctttttgag 2640  
 acaggttctc aatacttagc tcatcccaac ttgacccac tcttctcttg ccttagtcac 2700  
 cacaatgttt agtttataag catgctcac tatgcccggc tttaaaataa ctccaccata 2760  
 atccagcac tgaagtagac aaaagggagg atcgatgggg ctgactggcc acaagccgtg 2820  
 cttcaagttc aatgaagacc ctgtctcaag ggaataaggc acagaggata gagccatacg 2880  
 cctgacctcc tctctggcc tctacccagg cacatgtgca tacacacac acacacacac 2940  
 acacacacac acacacacac acagagagag agagagagag agagagagag agagagagag 3000  
 agagagagaa actttttctt cttttttttt aaaaatatta tttatttcat gtatatgagt 3060  
 acactgttgc tgtcttcaga cccccaaaa gaaggcatca gatccattg acaaatggtt 3120  
 gtgagccacc atgtggttgc tgggaattga actcaggacc tctggaaaag cagtcagtgc 3180  
 tcttaaccat ctcttcagcc ccacaaagaa acttttaatg agcaataat tgcttccaag 3240  
 taaatactac taatatattt ctaaccatac tatacaagga attattaag aacggataat 3300  
 aggagaataa aaaattataa gtcactttat aatcctatct aatccatcta gaacaaaaac 3360  
 actgtaataa tgcaaaagag cgcagtgcct agattaaata aataaaatgc agaccaataa 3420  
 gtaaaacttta tagcagcaca tggaaatgac gaaattccta acaaaaagct caagatgggc 3480  
 agtttatttta aagtgaataa caggagaat aaagcacaga aagatactca aaggcataga 3540  
 agttaacata ggggggctgg cgagatggct ctacaggtaa gagcaccoga ctgctcttcc 3600  
 aaaggtcctg agttcaaact ccagcaacca catggtggct cacaaccatc cgtaatgaga 3660  
 tttgactccc tcttctgggt tgtctgaaga cagctacagt gtacttaaca tataataaat 3720  
 aaataaaaact ttaaaaaaaa aattaaagaa gtttaacatag aagccactc aggacccac 3780  
 tcagtcttag agtatgacat tattatggac attaaaaaga gaaaattcag cagtagtbtg 3840  
 catgcactgc atatatcac aaatccttga gtttcatacc aaatgccttt agaccacttg 3900  
 tggctctgca aacctgtaat cctagcactt gtgaaaaggc cagctcagga actttgggaa 3960  
 ggtcatgaaa ctcttgcccc tccagaaggg agaggctaatt taacatttct cagaccacag 4020  
 ggcgggaacc gacctgcccc tggggacaga ctgttgcccc tttccagact agggaggtcc 4080  
 ttgtcacctc attccctaaa gaccaatcaa ccttagaaac cgtataaaaa ctagcgaagg 4140  
 ttgtgcttag ttgtgatgc tctattctgc tctgggacaa tccactaca ctggaacaa 4200  
 ggtaccaggg gtaacccctt ctcttcagg tctgggacaa tccactaca ctggaacaa 4260  
 aatttctctt ggctttttgc attgatcaca gctccacttc gtggtaagtt aagactccct 4320  
 ggagtcttac attggcaaat gcaggcaaaa gaatccgaaa ctcaaggta tctaaaacta 4380  
 catagcaagc atgtctgtag cctgggctcc atgagacctt gggggagggg cagagggaga 4440  
 ccgttcagaa gacagtcaag atgttgacag agcacaggca gcctggccac cagtgtgtc 4500  
 accagacatg ttaatgttgg aataaagcct caatcatgac tctcccagtt ttataattgg 4560  
 aaataagaaa ggaaagacta taggaacaac tgtgttcaga acactattta taatagcaaa 4620  
 gatctcagag taacccaaac ttctagacat tgaatttggg agatctcttg gcagcttatt 4680  
 ttgaaaactt tacaatgtta aatatgtaaa aacaaggaca gttttgtttt ttgtttgtt 4740  
 ttgttttgtt ttagggatat atattcatat atgtatatga atgaaaacc aaacttaaaa 4800  
 ttccccacta tgcttttaag gctttctgac aataacagaa agagaaatag agaatacata 4860  
 aaaactagtt ctgaaactat caataggctt gacactcttt agctgccagg agagctgaat 4920  
 ctgaacacag ggaacccac ccagaccccc aaatttggat tattgtttta ttttatcttt 4980  
 cccctacccc caagacaggg tttctctgctg tggctctggc tgtcctggaa ctccggagatc 5040

ctctgcctct	ctgcctctct	gcctctctct	ctctctgcct	ctctctgcct	ctctctctct	5100
ctctctctgc	ctctctctct	ctctctgccc	ctcgctgcct	ctctctgcct	ctctctgccc	5160
ctctctgccc	ctctctgccc	ctctctcttc	ccctctctgc	ctctctctct	gcccctctct	5220
gcctctctgc	ctctgcctcc	tgagtgtctg	gattttaaagg	catcagccat	cacttccagc	5280
ttcctttatc	attttaaaaa	gaatttccta	tgtgactact	gtattttaa	caccacacgg	5340
ccaatactcc	cccccaactc	ctoccaaactc	ccctctaccc	actcaaattc	ttatcttgta	5400
ttctttatca	ttattataca	tatgtgtata	tatgtgtgtg	tgtatatata	tataactat	5460
atactgctaa	tgagtaacat	ttagtggttat	tcattgttgc	atgttttcaa	tgtgctttcc	5520
aggaggctgg	ggggatggct	cagtgggcaa	aattctagct	gcacaagcct	aaggaccagg	5580
gttcagatcc	ccaatataaa	ggctggctgg	acatgggtgg	ttgcctatga	tactagcatg	5640
cttgctggaa	gcaaagacag	ggaatccctg	gagacttaga	atctcagaag	tgatctgggc	5700
tggacagact	agctgaactg	gccagctctg	ggttcatcaa	gaaaccctac	ctccataaca	5760
taaagtgtga	tggagaaagg	cacctaatgt	caacctcaaa	cccctacctg	catgtgcaca	5820
cacatacatc	cacaccacac	acacacacac	acacacacac	acacaccaca	cacacacaca	5880
cacaaataaa	taagtaaata	aataaaaatat	ttagctctcc	agaccaaata	ttggtgaaac	5940
ccatgcattt	gcatttgtgt	gtgtcctaca	aacactgaag	gttaagaagc	atgctcctta	6000
gtaattttat	ccagtttgc	gtttccagat	tgaaacacaga	ttctataggc	tacacagtgc	6060
taaattggatt	atgctcagat	acagattgaa	aaggatacag	attgaaaagg	gtcggggtct	6120
gggcccaggat	gacgggcca	ctatctttgc	ccgggcttgt	ccttcaggga	aggggttacag	6180
gattcaccac	tgggggtgtg	cctatctgct	gttaggacct	gaattgcctg	gagtgtttct	6240
agttcccaact	agttgttgaa	ctttaccttg	aacctctgct	cccagggaag	tcatcaggac	6300
tctgccatcc	ctggagtctc	tgcagagggt	gtttgaccaa	cagctttccc	caggccttcg	6360
cccacgacct	caggtaaagg	gtttggattt	ggaaagatgc	aattgctata	ggagggactc	6420
tgaaggcaga	cagacgcacc	gcctcctcac	tgtggctagt	ctaataataa	catcgcggtg	6480
gatggtgagg	atagactcca	tgcctttttg	tgaaggcatt	tcctggcatc	agctcctgac	6540
ttcagagttt	cacccatcag	acaaattgcc	tgggtgttga	ggaggagggtg	agcaggggcca	6600
ttcccatcat	ttctcctcag	aaatggaaag	gcaaggaaaa	catgaggttc	ttcagacact	6660
taatccctgg	gactgcaaaa	tgggtggtgcc	cctcctccac	agctgctcac	ggcggggcag	6720
gagatgaggg	ccgaatgaag	catagatcta	gctatttttt	ttttagtgc	ttcagtaaat	6780
ttaaaatcaa	ataagggaag	ggacctagat	ctttatgtta	tggcattgtt	aaaagtga	6840
actttagccc	aggggtgtgt	ggcgcacacc	tttaatccca	gcacttggga	ggcagaggca	6900
ggcggatttc	tgagtttgag	gccagcctgg	ttacacagat	gagttccagg	acagccaggg	6960
ctacacaggt	ttctctatac	agagaaacct	gtctcgaaaa	accaagaaaa	aaaagaaaga	7020
aaaagaaaga	aagaaagaaa	gaaagaaaga	aagaaagaaa	gaagaaagaa	aggaaaggaag	7080
gaaggaaagga	aggaagaaag	aggacaacat	gggtctagggg	tcagagagca	gatctccaaa	7140
aacaccaaca	atgccttgct	gtaaatgtat	gtcggttgatt	tggggatgtt	ggcctccagc	7200
tcaccatttc	ctgccttagc	ttccaaagtg	ctaggattat	aggcttgagc	caacacatct	7260
ggcttagcgt	tattgtgtgt	ggaaggggag	tgtgagtgt	gctcctgtgt	ttggtactta	7320
tatatgaata	tatgtatata	cgcattgtacg	catacttgca	tgtgaaggcc	agaggccaat	7380
gtcagctgct	ctcatcttat	cctttttatt	acattgtatt	tatttgtttg	tttgtttgtt	7440
tgtttcacog	tacgcatgca	gccactcatg	agcatgtaac	agcacaggta	tgaaggtaga	7500
cttgaggag	tcagttctct	ccttctgtca	cttgagttcc	aggaccacac	tccagccccc	7560
aggcctgggc	tgtaaagagag	ccatcttact	ggctcctctac	tttgtcttct	gagatagcat	7620
ctagactcac	ggaacctgga	gctcatctag	atttacattg	gctggccagc	tgatgcattt	7680
taagggtcaaa	tcttcattcc	atccctaccc	cacttccact	cccagtgtcg	gagttcggga	7740
cacctgccac	caagcccagt	ttttcctgga	tgcagaagct	ccaaactcag	gttcccatgt	7800
tcgcatggca	ggcacathtt	cagttaagcc	ttccccccag	ctcctttacc	ctgggtctctg	7860
aatggggggg	aggctataaa	tcaggctgct	ctcagacatt	aggtaggaaa	tagaccatat	7920
acatgaggaa	agatattcac	ctgccccatg	ggtagcagga	agtgatgtcc	aactcctctt	7980
ttgcttatca	ggagaaatgc	tgactactac	ctctggtaat	tttgatgttg	ggaggaaacag	8040
ggacattcat	aggaccccat	tcctcgctgg	tgagagtggga	gacaggtttt	ctgaagggca	8100
ggagatctgt	gtagaaaaga	tggatgctgt	tttctgaagg	gaaatggagg	tagagtcgac	8160
ctgggagaga	gggaagggtg	gggatgggtg	ggaggaatga	aaggaagaga	gatggcagtt	8220
gggatgtatt	gtataagaga	agaatcagaa	agaaaaaaag	aaaagctacc	tgcacccttc	8280
aagtgttcc	ctgtgtggga	ggctgtctca	gggactacat	gggcaccgag	aggcatcagt	8340
gagggtaggt	acttgatgtt	gtgtccctga	aaacaaggac	aggaaatctg	ctgcatggcc	8400
taagatggca	aatgtggca	caatcaagta	aggcccagga	ttctgtctgt	ggtgcagacc	8460
tgtgtagaa	tgagctccca	gcattccac	ttgtgtgtgt	gagacagcat	gttgcagagc	8520
catgtgagga	tgagggtcca	ggccaggagg	atgtcaaccc	accaccatgt	agccagtggg	8580
ctgggggagc	ttggggccac	caaggagctt	gagcagactg	acagtgggtt	atgtacacaa	8640

```

gtgggcgtgt cacacaaccg tgcaacacag agaaaatccc tgtgatgaca acttctaaac 8700
caccctgagg caaaaggagt agacagggga ttagagccta gcatattgga gtcgagtggc 8760
catgcagctc ttggaagcgt gaggaaggaa atttcctgga aggatagggt gtcttcctag 8820
cagcctcgtc aatagatgtc aatgtatgag gtagtacctg ctacaatcct gcttcttcag 8880
aagactgagg caggggggatt acttgaaccc agaagttcta ggccagtatg gacaacatag 8940
caagagtctg atttaaaaaa aaaaaaaaag taaagaggga aaccaaatag gtgacgtgcc 9000
acactagtgt ctctctgctc caagggtcct gggcacatga gcttgccttag tgccagaaaa 9060
gtcagaggag gagagggcag acagagaccc tcgtctccac ctctcttgac tgactaatgg 9120
ggctggatat aatctgtttt acaaaaggac agcttttcag agctgtttct atctaagggt 9180
gctctgaata gccatctcga aatatgccag agaagaatat ttagaggcgg cctatttttg 9240
tctcccacaa agatttcaca gggaaaatgt atttgtgttc tatttataca actaaaaata 9300
tgcacagacc cggggaaact ggctctttgc tgcttttgaa gtgaaggatg gggttaattc 9360
taagaaagta aaagcaaagtg tagtgccagg accacggatg ctgccagaca ccagcgttta 9420
agtggcttga atggaagagc acaccccaag tatttgaagt aggtgaggag agagaccagg 9480
tctccagagt tgggcctgca gtggccaggg taagtccaag gggcagtatg acgcagaaca 9540
gagtggcaac ctctaccagt agtagaatc aggtctcatt cctaacctcc cataaagcag 9600
agaatattgc actctctttc tctctctctc tctctctctc tctctctctc tctctctctc 9660
tctctctctc tctaactcac agagatcctc ctgcctctgc ttcctcattg ctggaaataa 9720
aagcatgtgc caccacatgc agctctcctt gtttggagag agaaaggcag acagacagac 9780
agacagacag acagacagag tataatgtgt gcagatgtcc acagaactcc gaagagggtg 9840
tgggaccccc agaaaccgga gttctaggca gttgtaagct gtccagtggg tgttgccaac 9900
tgaactcaag tcctctggaa aactggaaag tactcttaac cactaagcta tctgttagtc 9960
ccccagaat gtcttatctt gataggcctt catagctcac agtccagggg gctgactgct 10020
acaaggggta gaggaagaa aactggcccc aggcctgagt tcagagtgc atctccagag 10080
ttcttctctc ccttgccgtg tagtaatttc ctacctaca cgaggagaaa aggaacagat 10140
atgtccagta tgcctggcat cttgaaaggg cactgactct tgccgctgta gagccccctt 10200
ccttggggag tgcagagaag tgctgctaga gaggttcaa agaaacacaa cagtcaaaat 10260
agttgctggc caagggaggg catgtgccct gtatcgggcc tgcaaagccc tttcaacagt 10320
tagccaagg ccatgtttga cagtacaagc ctgtaagccc atgcataagc aagggtctgg 10380
ataaagggct actgttcaag ttagttatat acacatcaag tttgttcac ttacatcctt 10440
aggtaagggt gggtagcttg ttagctccac ctctccagca ggaaagcatg tccacggtaa 10500
ggtagatact gtagagtta gctttgtttt gaccttctt tcttcgttgc ttgggattgt 10560
acttgggaac ttcactcatg cagcccatgt ttcagccttg cataaatcta taaatatatg 10620
gttggagtca ggtccagcat atgtgggtac cagtcccagt ctggtatagc aatttattag 10680
ctatttgtct ttagacaagt aactataggg ttctgagctg taatggccac caaacagggtc 10740
ctttaagtat ctttattggc caaacaccag ggttaaaaca ctgaacacag tatggtggtt 10800
tcaactgtca tggagtcac attctagcag gggaatttta cacacaactc acaaatgata 10860
cagaccacag cagggtgatt tgacagagac aacgaagggc ttgctgagc tgcatgagtc 10920
cgggagcagc tcttgatggg atctgagatc tagatggcag agcgcccttc ctatgtaaa 10980
tccaagcctc tggttactcc tcagcaagag tagctggagg aagaccagaa gtgaggagga 11040
aaaggctcag agtggagcat ctaagcctgc agcacctcca gagacaggct ttggagcttg 11100
cctgtgggtc agcctactgt gggagagac cagtgagggt ttaagctgt ggggatgggt 11160
atattttgga aggcgaacaa gaacaccagc agctcaggca atgtaggagg gcgggaagg 11220
gatggatgat caggacggc ccagagatt ttaaggactg ttaggtggg gtgagaagca 11280
ctgtgggtga ggtggagtta gaggggaagaa gtgaagatct atttttttgg ccatgaggaa 11340
tttgggtgac acacacacac atacatacac acccatcat gtcttgtct acagacgggg 11400
tgataatggt actggaaggt agaaggtgga aagaggacca gcaaggaagg tatccagtgc 11460
ccacccca gctcaccctc agacagacct tgttctcatt ttcattacc aggtgcccg 11520
agaggccagt cccatcacc

```

&lt;210&gt; 6

&lt;211&gt; 22

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: Synthetic  
Primer

<400> 6  
atctttgccc ggggcttgct ct 22

<210> 7  
<211> 50  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
Primer

<400> 7  
gtctgggcca ggatgacggg ccaactatct ttgcccgggc ttgtccttca 50

<210> 8  
<211> 12  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
Primer

<400> 8  
cactagaagg tt 12

<210> 9  
<211> 13  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
Primer

<400> 9  
gtgcgtgggc cag 13

<210> 10  
<211> 10  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
Primer

<400> 10  
tcaggagag 10

<210> 11  
<211> 12  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
Primer

<400> 11  
gtgcgcctat ct 12

<210> 12  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
Primer

<400> 12  
tactcctcag caagagtagc tgg 23

<210> 13  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
Primer

<400> 13  
gctgaacttg tggccgttta cgt 23

<210> 14  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
Primer

<400> 14  
cttctatagc cttcccaagc c 21

<210> 15  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
Primer

<400> 15  
ctcgtaggtc tcacaggaag 20



<210> 16  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
Primer

<400> 16  
cctgctggat tacattaaag cactg 25

<210> 17  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
Primer

<400> 17  
gtcaagggca tatccaacaa caaac 25

<210> 18  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
Primer

<400> 18  
ggcgttctct ttggaaaggt gttc 24

<210> 19  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
Primer

<400> 19  
ctcgaaccac atccttctct 20

<210> 20  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic

## Primer

<400> 20  
ttgctgcacg agggctcaga atc 23

<210> 21  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
Primer

<400> 21  
tccgattctc atgctctggc ttg 23

<210> 22  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
Primer

<400> 22  
cagccctgtc ttgcccacgt ttg 23

<210> 23  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
Primer

<400> 23  
tccactggat tcatcccgct ctg 23

<210> 24  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
Primer

<400> 24  
cttcctcttg caacagagaa ccc 23

<210> 25

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
Primer

<400> 25

actcaacgtc cactttgaga tgc

23

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US2005/008977

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 7 C12N9/12 C12N15/85

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the International search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, Sequence Search, EMBASE, PAJ, WPI Data, CHEM ABS Data

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	-/-	

☒ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

\* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- \*G\* document member of the same patent family

Date of the actual completion of the international search

20 July 2005

Date of mailing of the International search report

05/08/2005

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Mundel, C

## INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US2005/008977

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>TU ZHENG ET AL: "Embryonic and hematopoietic stem cells express a novel SH2-containing inositol 5'-phosphatase isoform that partners with the Grb2 adapter protein"</p> <p>BLOOD, vol. 98, no. 7, 1 October 2001 (2001-10-01), pages 2028-2038, XP002336919 ISSN: 0006-4971 cited in the application abstract page 2028, right-hand column, line 5 - line 14 figure 3 page 2033, left-hand column, line 6 - line 13 page 2034, right-hand column, line 55 - line 59 page 2037, left-hand column, line 8 - line 13</p>	1-76
X	<p>SLY LAURA M ET AL: "SHIP, SHIP2, and PTEN activities are regulated in vivo by modulation of their protein levels: SHIP is up-regulated in macrophages and mast cells by lipopolysaccharide."</p> <p>EXPERIMENTAL HEMATOLOGY (NEW YORK), vol. 31, no. 12, December 2003 (2003-12), pages 1170-1181, XP002336920 ISSN: 0301-472X cited in the application abstract page 1173, right-hand column, line 23 - line 43</p>	1-76
X	<p>ROHRSCHEIDER LARRY R ET AL: "Structure, function, and biology of SHIP proteins"</p> <p>GENES AND DEVELOPMENT, vol. 14, no. 5, 1 March 2000 (2000-03-01), pages 505-520, XP002336921 ISSN: 0890-9369 cited in the application page 506, right-hand column, line 19 - line 26 page 508, right-hand column, line 19 - line 24</p>	1-76

-/--

## INTERNATIONAL SEARCH REPORT

International Application No.

PCT/US2005/008977

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KAVANAUGH W M ET AL: "MULTIPLE FORMS OF AN INOSITOL POLYPHOSPHATE 5-PHOSPHATASE FORM SIGNALING COMPLEXES WITH SHC AND GRB2" CURRENT BIOLOGY, CURRENT SCIENCE,, GB, vol. 6, no. 4, 1 April 1996 (1996-04-01), pages 438-445, XP000197368 ISSN: 0960-9822 cited in the application the whole document	1-76
A	WOLF I ET AL: "Cloning of the genomic locus of mouse SH2 containing inositol 5-phosphatase (SHIP) and a novel 110-kDa splice isoform, SHIPdelta." GENOMICS. 1 OCT 2000, vol. 69, no. 1, 1 October 2000 (2000-10-01), pages 104-112, XP004437818 ISSN: 0888-7543 cited in the application the whole document	1-76